

METHODS AND COMPOSITIONS FOR SCREENING FOR
MODULATORS OF IgE SYNTHESIS, SECRETION AND
SWITCH REARRANGEMENT

FIELD OF THE INVENTION

The invention relates to methods and compositions useful in screening for modulators of IgE synthesis, secretion and switch rearrangement.

BACKGROUND OF THE INVENTION

15 Immunoglobulins must bind to a vast array of foreign molecules and thus exist in many forms. The sequence of the variable (V) region of immunoglobulin molecules varies tremendously, conferring virtually unlimited capacity to bind antigens. The constant (C) region comes in five different varieties: α , δ , ϵ , γ and μ , providing five different isotypes: IgA, IgD, IgE, IgG and IgM, each of which performs a different set of functions. B cells initially produce only IgM and IgD, and must be activated or induced to produce the other isoforms, such as IgE.

20 The course of IgE production starts with the activation of B cells. Upon activation with an antigen, B cells follow one of two differentiation pathways: they may differentiate directly into plasma cells, which are basically antibody-secreting factories, or they may give rise to germinal centers, specialized structures within lymphoid organs. In the latter, successive

rounds of mutation of the V region genes is followed by expression of the gene products on the cell surface, with selection of the cells on the basis of the affinity of the mutated immunoglobulins against the antigen.

5 In both pathways of antigen-induced B cell differentiation, isotype switching occurs in which the C region of the immunoglobulin heavy chain changes from the joint expression of IgM and IgD on naive B cells to expression of one of the downstream isotypes such as IgE. This switching involves the replacement of upstream C regions with a downstream C region that has biologically distinct effector functions without changing the structure of the variable portion and, hence, its specificity. For IgE switching, a deletional rearrangement of the Ig heavy chain gene locus occurs, a rearrangement that joins the switch region of the μ gene, $S\mu$, with the corresponding region of the ϵ gene, SE . This switching is minimally induced by IL-4 or IL-13, which initiates transcription through the SE region, resulting in the synthesis of germ-line (or "sterile") ϵ transcripts; that is, transcripts of the unarranged C_ϵ heavy genes. This IL-4 induced transcription is inhibited by IFN- γ , IFN- α , and TGF- β . A second signal, normally delivered by T cells, is required for actual switch recombination leading to IgE production. The T cell signal may be replaced by monoclonal antibodies to CD40, Epstein-Barr viral infection, or hydrocortisone.

10 20 Recently, the mechanism of class switch recombination has been explained by an accessibility model, wherein the specificity of the switch gene rearrangement is determined by the modulation of switch region accessibility; that is, the opening up of the chromatin in certain areas, allowing the required protein/enzyme complexes access to the genes.

15 25 IgE antibodies are crucial immune mediators of allergic reactions, and have been shown to be responsible for the induction and maintenance of allergic symptoms. For example, the introduction of anti-IgE antibodies has been shown to interfere with IgE function, thus working to alleviate allergic symptoms. See Jardieu, Current Op. Immunol. 7:779-782 (1995), Shields et al., Int. Arch. Allergy. Immunol. 107:308-312 (1995).

Accordingly, it is an object of the invention to provide compositions and methods useful in screening for modulators of IgE production, in particular for modulators of switch rearrangement.

5

SUMMARY OF THE INVENTION

In accordance with the objects outlined above, the present invention provides methods of screening for bioactive agents capable of inhibiting an IL-4 inducible ϵ promoter. The method comprises combining a candidate bioactive agent and a cell comprising a fusion nucleic acid. The fusion nucleic acid comprises an IL-4 inducible ϵ promoter, and a reporter gene. The promoter is then induced with IL-4 (or IL-13), and the presence or absence of the reporter protein is detected. Generally, the absence of the reporter protein indicates that the agent inhibits the IL-4 inducible ϵ promoter. The fusion nucleic acid may comprise an exogenous IL-4 inducible ϵ promoter, or an endogenous IL-4 inducible ϵ promoter. Preferred embodiments utilize the use of retroviral vectors to introduce the candidate bioactive agents.

10
15
20
25

In an additional aspect, the present invention provides cell lines for screening. Either CA-46 and MC-116 cell lines are included, and further comprise fusion nucleic acids comprising an IL-4 inducible ϵ promoter, and a reporter gene.

In a further aspect, the present invention provides methods of screening for bioactive agents capable of modulating IgE production. The method comprises combining a candidate bioactive agent and a cell capable of expressing IgE and determining the amount of IgE produced in the cell. Generally, a change in the amount of IgE as compared to the amount produced in the absence of the candidate agent indicates that the agent modulates IgE production. The cell can further comprise a IgE fusion protein comprises the ϵ heavy chain, and a fluorescent protein.

In an additional aspect, the invention provides methods of screening for bioactive agents capable of inhibiting a promoter of interest. The method comprises combining a candidate bioactive agent and a cell comprising a fusion nucleic acid. The fusion nucleic acid comprises a promoter of interest and a reporter gene comprising a death gene that is activated by the introduction of a ligand. The promoter is optionally induced, and the ligand is introduced to the cell. The presence of the cell is then detected, wherein the presence of the cell indicates that the agent inhibits the promoter.

In a further aspect, the invention provides compositions comprising a test vector and a reporter vector. The test vector comprises a first selection gene, and a fusion gene comprising a first sequence encoding a transcriptional activation domain, and a second sequence encoding a test protein. The reporter vector comprises a first detectable gene, and all or part of the switch ϵ sequence, which upon binding of the transcriptional activation domain due to a protein-nucleic acid interaction between the test protein and the switch ϵ sequence, will activate transcription of the first detectable gene. Methods utilizing these compositions are also provided; the methods comprise providing a host cell comprising the composition, and subjecting the host cell to conditions under which the fusion gene is expressed to produce a fusion protein. A protein-nucleic acid interaction between the fusion protein and the switch ϵ sequence is then detected.

20

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B depict the germline ϵ locus and sequence. Fig. 1A depicts the sequence

25 of the human IL-4 inducible ϵ promoter. Fig. 1B depicts the organization of the germline ϵ locus.

Figures 2A and 2B depict the regions (2A) and sequences (2B and 2C) of the switch ϵ (S ϵ) region that are used in methods of screening for proteins that interact with the S ϵ region, as described below.

30

Figure 3 shows a schematic of the yeast one-hybrid system used to identify proteins that bind to the S ϵ region.

Figure 4 depicts the IL-4 induction of germline ϵ mRNA in three IgM $^+$ B cell lines, CA-46,

5 MC-116 and DND39. The cells were incubated for 48 hours in 300 U/ml of hIL-4. RT-PCR was performed using primers specific for the germline ϵ exon and the 5'-end of the ϵ CH1 exon (predicted size is ~200 bp).

10 Figures 5A, 5B, 5C and 5D depict two general approaches to generate germline ϵ promoter knock-in reporter cell lines. Figure 5A shows the organization of this region in vivo. Figures 5B and 5C depict two possible knock in constructs. The IL-4 inducible IgM $^+$ B cell lines are transfected with one or both of these constructs. Under the influence of IL-4, GFP and/or BFP positive clones are isolated by FACS. Homologous recombination can be confirmed by PCR and/or Southern blot hybridization. Figure 5D depicts an alternate construct. In this 15 embodiment, the IL-4 inducible IgM $^+$ B cell lines are transfected with the 5D construct and selected with G418. Survivors are sorted for the lack of the 3' BFP expression (deleted during homologous recombination). RT-PCR is performed to confirm homologous recombination. Those clones are transfected with cre to remove the neomycin resistance gene.

20 Figure 6 depicts a preferred vector for introducing a peptide library into cell lines containing knock-in reporter genes under the control of the IL-4 inducible ϵ promoter. CRU5 is a modified LTR; Naviaux, et al., "The pCL Vector System: Rapid Production of Helper-Free, High-Titer, Recombinant Retroviruses," *Journal of Virology*, 70(8):5701-5705 (1996); LTR = long terminal repeat; Ψ^+ = packaging signal; localization signal = nuclear, cell membrane, etc.; MCS = multiple cloning site; IRES = internal ribosome entry site; 2a = self-cleaving peptide. All the components are cassetted for flexibility.

25 Figure 7 depicts a general schematic of the generation of the primary peptide libraries in 30 retroviruses.

Figures 8A and 8B depict constructs useful in generating ϵ heavy chain knock-in cell lines. Figure 8A depicts the wild-type organization. Figure 8B depicts a representative construct to produce a GFP knock-in. S = secretory exon; GFP = green fluorescent protein; BFP = blue fluorescent protein; Neo^r = neomycin resistance gene; VDJ = V region exon; CH1, 2, 3, 4 = constant region domain exons; M1, M2 = membrane exons; HSV-TK = Herpes Simplex Virus - thymidine kinase.

Figures 9A and 9B depict constructs useful in the invention. Figure 9A shows a reporter construct useful to create an IL-4 inducible ϵ promoter reporter cell line. CRU5 = hCMV promoter plus R and U5 regions of LTR; BGH poly A = bovine growth hormone polyadenylation signal; SIN = self-inactivating LTR. Figure 9B shows a library construct.

Figures 10A and 10B depict a schematic of the screen for candidate agents of the germline ϵ promoter. Figure 10A: the experimental schematic. Figure 10B depicts the survival construct useful in the screen. Position 1 can be a number of different genes, including a FAS chimeric receptor outlined herein (including extracellular mouse Fas receptor or mouse CD8 receptor coupled with the human transmembrane and cytoplasmic Fas receptor), HSV-TK, p450 2B1 and p21 peptide.

Figures 11A, 11B and 11C depict preferred vectors and their sequences.

Figures 12A, 12B and 12C depict a construct useful in the present invention, comprising the a Fas survival construct (i.e. the use of a death gene). The sequence is of the inducible ϵ promoter-chimeric Fas-IRES-hygromycin-bovine growth hormone poly A tail that is put into the C12s vector backwards to that no leaky transcription happens through the cmv promoter.

Figures 13A, 13B and 13C depict a construct useful in the present invention, comprising the a Fas survival construct (i.e. the use of a death gene). The sequence is of the inducible ϵ promoter-chimeric Fas (either CD8 or mLyt2)-IRES-hygromycin-bovine growth hormone

poly A tail that is put into the C12s vector backwards to that no leaky transcription happens through the cmv promoter.

DETAILED DESCRIPTION OF THE INVENTION

5

The present invention provides compositions and methods useful in screening for modulators, particularly inhibitors, of the production of IgE antibodies. In particular, assay methodologies are provided that are amenable to high-throughput screening strategies, such that large numbers of potential drugs may be screened rapidly and efficiently. Generally, traditional treatments for IgE suppression are based on regulation of the system after IgE has been made, for example using anti-IgE antibodies or anti-histamines, to modulate the IgE-mediated response resulting in mast cell degranulation. In some cases, drugs are known that generally downregulate IgE production or that inhibit switching but not induction of germline transcripts (see for example Loh et al., *J. Allerg. Clin. Immunol.* 97(5):1141 (1996)).

10

In contrast, the present invention provides several related techniques that may be used to screen for upstream modulators of IgE production, to prevent the production of IgE and thus reduce or eliminate the allergic response. For example, an early step in the Ig switch is the production of sterile ϵ transcripts in response to IL-4. It is also appreciated that blockage of 20 the production of membrane bound IgE may induce programmed cell death (PCD). By interfering at this step, highly efficient, rapid and prolonged inhibition of the allergic response may occur. In addition, these techniques allow individual cell assessment and thus are useful for high-throughput screening strategies, for example those that utilize fluorescence activated cell sorting (FACS) techniques, and thus allow screening of large numbers of compounds for 25 their effects on IgE production.

In a preferred embodiment, the invention relates to methods that rely on reporter genes fused to IgE promoters, such as the IL-4 inducible ϵ promoter that starts a cascade that ultimately results in IgE production. Using novel reporter constructs, screening for modulators of this

promoter system may be done. Thus the invention provides a number of different constructs that allow for screening for antagonists and agonists of these promoters.

5 In a preferred embodiment, the invention provides methods of screening for bioactive agents capable of modulating, particularly inhibiting, an IL-4 inducible ϵ promoter. By "an IL-4 inducible promoter" herein is meant a nucleic acid promoter that is induced by IL-4, putatively by binding an unknown IL-4 induced DNA binding protein that results in induction of the promoter; that is, the introduction of IL-4 causes the pronounced activation of a particular DNA binding protein that then binds to the IL-4 inducible promoter segment and induces transcription. The sequence of the human IL-4 inducible promoter is shown in Figure 1, and as will be appreciated by those in the art, derivatives or mutant promoters are included within this definition. Particularly included within the definition of an IL-4 inducible promoter are fragments or deletions of the sequence shown in Figure 1. As is known in the art, the IL-4 inducible promoter is also inducible by IL-13. By "modulating an IL-4 inducible promoter" herein is meant either an increase or a decrease (inhibition) of promoter activity, for example as measured by the presence or quantification of transcripts or of translation products. By "inhibiting an IL-4 inducible promoter" herein is meant a decrease in promoter activity, with changes of at least about 50% being preferred, and at least about 90% being particularly preferred.

10

15

20

The methods comprise combining a candidate bioactive agent and a cell or a population of cells comprising a fusion nucleic acid. The cell or cells comprise a fusion nucleic acid. In a preferred embodiment, the fusion nucleic acid comprises an IL-4 inducible ϵ promoter and at least a first reporter gene. The IL-4 inducible ϵ promoter is as described herein, for example SEQ ID NO:1, or derivatives thereof, and may be either an endogenous or exogenous IL-4 inducible ϵ promoter, as is more fully described below.

25

30 By "reporter gene" or "selection gene" herein is meant a gene that by its presence in a cell (i.e. upon expression) can allow the cell to be distinguished from a cell that does not contain the reporter gene. Reporter genes can be classified into several different types, including

5 detection genes, survival genes, death genes and cell cycle genes. It may be the nucleic acid or the protein expression product that causes the effect. As is more fully outlined below, additional components, such as substrates, ligands, etc., may be additionally added to allow selection or sorting on the basis of the reporter gene.

5

10 15 20

15

20

In a preferred embodiment, the reporter gene encodes a protein that can be used as a direct label, i.e. a detection gene, for sorting the cells, i.e. for cell enrichment by FACS. In this embodiment, the protein product of the reporter gene itself can serve to distinguish cells that are expressing the reporter gene. In this embodiment, suitable reporter genes include those encoding green fluorescent protein (GFP; Chalfie, et al., "Green Fluorescent Protein as a Marker for Gene Expression," *Science* 263(5148):802-805 (Feb 11, 1994); and EGFP; Clontech - Genbank Accession Number U55762), blue fluorescent protein (BFP; 1. Quantum Biotechnologies, Inc. 1801 de Maisonneuve Blvd. West, 8th Floor, Montreal (Quebec) Canada H3H 1J9; 2. Stauber, R. H. *Biotechniques* 24(3):462-471 (1998); 3. Heim, R. and Tsien, R. Y. *Curr. Biol.* 6:178-182 (1996)), enhanced yellow fluorescent protein (EYFP; 1. Clontech Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303), luciferase (Ichiki, et al.), and β -galactosidase (Nolan, et al., "Fluorescence-Activated Cell Analysis and Sorting of Viable Mammalian Cells Based on Beta-D-galactosidase Activity After Transduction of Escherichia Coli LacZ," *Proc Natl Acad Sci USA* 85(8):2603-2607 (Apr 1988)).

25

25

30

Alternatively, the reporter gene encodes a protein that will bind a label that can be used as the basis of the cell enrichment (sorting); i.e. the reporter gene serves as an indirect label or detection gene. In this embodiment, the reporter gene should encode a cell-surface protein. For example, the reporter gene may be any cell-surface protein not normally expressed on the surface of the cell, such that secondary binding agents could serve to distinguish cells that contain the reporter gene from those that do not. Alternatively, albeit non-preferably, reporters comprising normally expressed cell-surface proteins could be used, and differences between cells containing the reporter construct and those without could be determined. Thus, secondary binding agents bind to the reporter protein. These secondary binding agents are

preferably labelled, for example with fluors, and can be antibodies, haptens, etc. For example, fluorescently labeled antibodies to the reporter gene can be used as the label. Similarly, membrane-tethered streptavidin could serve as a reporter gene, and fluorescently-labeled biotin could be used as the label, i.e. the secondary binding agent. Alternatively, the 5 secondary binding agents need not be labeled as long as the secondary binding agent can be used to distinguish the cells containing the construct; for example, the secondary binding agents may be used in a column, and the cells passed through, such that the expression of the reporter gene results in the cell being bound to the column, and a lack of the reporter gene (i.e. inhibition), results in the cells not being retained on the column. Other suitable reporter 10 proteins/secondary labels include, but are not limited to, antigens and antibodies, enzymes and substrates (or inhibitors), etc.

100
90
80
70
60
50
40
30
20
10
0

In a preferred embodiment, the reporter gene is a survival gene that serves to provide a nucleic acid (or encode a protein) without which the cell cannot survive, such as drug 150 resistant genes. In this embodiment, the assays may rely on clonal or pooled populations of cells, since if inhibitors of the promoter are found, the cells will die, necessitating a clonal population in order to determine the candidate agent.

In a preferred embodiment, the reporter gene is a cell cycle gene, that is, a gene that causes 20 alterations in the cell cycle. For example, p21 protein its ligand (a collection of three proteins; see Harper, et al., "The p21 Cdk-Interacting Protein Cip1 Is a Potent Inhibitor of G1 Cyclin-Dependent Kinases," *Cell* 75:805-816 (November 19, 1993)), which does not cause death, but causes cell-cycle arrest, such that cells containing inhibited IL-4 inducible promoters grow out much more quickly, allowing detection on this basis. As will be 25 appreciated by those in the art, it is also possible to configure the system such that the cells containing the inhibited IL-4 inducible promoters do not grow out, and thus can be selected on this basis as well.

In a preferred embodiment, the reporter gene is a death gene that provides a nucleic acid that 30 encodes a protein that causes the cells to die. Death genes fall into two basic categories:

death genes that encode death proteins that require a death ligand to kill the cells, and death genes that encode death proteins that kill cells as a result of high expression within the cell, and do not require the addition of any death ligand. It is preferable that cell death requires a two-step process: the expression of the death gene and induction of the death phenotype with 5 a signal or ligand, such that the cells may be grown up expressing the death gene, and then induced to die. A number of death genes/ligand pairs are known, including, but not limited to, the Fas receptor and Fas ligand (Bodmer, et al., "Characterization of Fas," *J Biol Chem* 272(30):18827-18833 (Jul 25, 1997); muFAS, Gonzalez-Cuadrado, et al., "Agonistic anti-Fas Antibodies Induce Glomerular Cell Apoptosis in Mice In Vivo," *Kidney Int* 51(6):1739-1746 10 (Jun 1997); Muruva, et al., *Hum Gene Ther*, 8(8):955 (May 1997)), (or anti-Fas receptor antibodies); p450 and cyclophosphamide (Chen, et al., "Potentiation of Cytochrome P450/Cyclophosphamide-Based Cancer Gene Therapy By Coexpression of the P450 Reductase Gene," *Cancer Res* 57(21):4830-4837 (Nov 1 1997)); thymidine kinase and 15 gangcylovir (Stone, R., "Molecular 'Surgery' For Brain Tumors," 256(5063):1513 (June 12, 1992)), tumor necrosis factor (TNF) receptor and TNF. Alternatively, the death gene need not require a ligand, and death results from high expression of the gene; for example, the overexpression of a number of programmed cell death (PCD) proteins are known to cause cell 20 death, including, but not limited to, caspases, bax, TRADD, FADD, SCK, MEK, etc.

25 As will be appreciated by those in the art, the use of the death genes in the manner described herein, particularly in two-step applications, allows general and high-throughput screening for inhibitors of other promoters, in addition to the IL-4 inducible ϵ promoters described herein. Thus, the present invention provides fusion nucleic acids comprising a promoter of interest operably linked to a death gene for use in screening methods. The promoter of interest can be either a constitutive promoter or an inducible promoter, such as the IL-4 inducible ϵ promoter. As will be appreciated by those in the art, any number of possible promoters could be used. Suitable promoters of interest include, but are not limited to, inducible promoters such as IL-4 ϵ promoter, promoters that are induced by cytokines or growth factors such as the interferon responsive factors 1 to 4, NFkB (Fiering, et al., "Single Cell Assay of a 30 Transcription Factor Reveals a Threshold in Transcription Activated By Signals Emanating

From the T-Cell Antigen Receptor,” *Genes Dev* 4(10):1823-1834 (Oct 1990)), etc. When inducible promoters are used in this embodiment, suitable cell types are those that can be induced by the appropriate inducer, as will be appreciated by those in the art.

5 Preferred embodiments fall into one of three configurations. In a preferred embodiment, the promoter of interest is a constitutive promoter, and it is hooked to a death gene that requires the presence of a ligand, such as Fas or TNF. Thus, the cells can be grown up and the presence of the death gene verified due to the constitutive promoter. This is generally done by hooking the death gene up to a detection gene such as GFP or BFP, etc., using either an IRES or a protease cleavage site as is outlined below; thus, the presence of the detection gene means the death gene is also present. Verification of the presence of the death gene is preferred to keep the levels of false positives low; that is, cells that survive the screen should be due to the presence of an inhibitor of the promoter rather than a lack of the death gene.

10 Once the cells have been enriched for those containing the death gene, the candidate agents can be added (and their presence verified as well), followed by induction in the presence of IL-4, and finally by addition of the death ligand. Thus, the cell population is enriched for those cells that have an agent that inhibits the promoter and thus does not produce the death protein, i.e. those that survive.

20 Alternatively, a preferred embodiment utilizes fusion nucleic acids comprising promoters of interest that are inducible (such as the IL-4 ϵ promoter), and hooked to a death gene that requires a death ligand. The presence of the death gene is verified by inducing the promoter, causing the death gene (and preferably a detection gene) to be made. The candidate agents 25 and death ligands are then introduced in the presence of their appropriate inducer, and the population is enriched for those cells that survive, i.e. contain an agent that inhibits the promoter and thus does not produce the death protein.

When death genes that require ligands are used, i.e. for “two step” processes, preferred 30 embodiments utilize chimeric death genes, i.e. chimeric death receptor genes. These chimeric

death receptors comprise the extracellular domain of a ligand-activated multimerizing receptor and the endogeneous cytosolic domain of a death receptor gene, such as Fas or TNF. This is done to avoid endogeneous activation of the death gene. The mechanism of Fas-induced cell death involves the introduction of the Fas ligand, which can bind two monomeric
5 Fas receptors, causing the multimerization of the receptor, which activates the receptor and leads to secondary signalling resulting in caspase activation and PCD. However, as will be appreciated by those in the art, it is possible to substitute the extracellular portion of the death receptor with the extracellular portion of another ligand-activated multimerizing receptor, such that a completely different signal activates the cell to die. There are a number of known
10 ligand-activated dimerizing receptors, including, but not limited to, the CD8 receptor, erythropoietin receptor, thrombopoietin receptor, growth hormone receptor, Fas receptor, platelet derived growth hormone receptor, epidermal growth factor receptor, leptin receptor, and a variety of interleukin receptors (including, but not limited to, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-11, IL-12, IL-13, IL-15, and IL-17; although the use of the IL-4 and IL-13 receptors are not preferred, since these can be used to induce the promoter and thus does not provide a "two step" death process), low-density lipoprotein receptor, prolactin receptor, and transferrin receptor.

15

20 In a preferred embodiment, chimeric Fas receptor genes are made. The exact combination will depend on the cell type used and the receptors normally produced by these cells. For example, when using human cells or cell lines, a non-human extracellular domain and a human cytosolic domain are preferred, to prevent endogeneous induction of the death gene. For example, a preferred embodiment utilizes human cells, a murine extracellular Fas
25 receptor domain and a human cytosolic domain, such that the endogeneous human Fas ligand will not activate the murine domain. Alternatively, human extracellular domains may be used when the cells used do not endogeneously produce the ligand; for example, the human EPO extracellular domain may be used when the cells do not endogeneously produce EPO. (Kawaguchi, et al., *Cancer Lett.*, 116(1):53 (1997); Takebayashi, et al., *Cancer Res.*, 56(18):4164 (1996); Rudert, et al., *Biochem Biophys Res Commun.*, 204(3):1102 (1994);
30 Rudert, et al., *DNA Cell Biol.*, 16(2):197 (1997); Takahasi, et al., *J Biol Chem.*

271(29):17555 (1996); Adam, et al., *J Biol Chem.*, 268(26):19882 (1993); Mares, et al., *Growth Factors*, 6(2):93 (1992); Seedorf, et al., *J Biol Chem.*, 266(19):12424 (1991); Heidaran, et al., *J Biol Chem.*, 265(31):18741 (1990); Okuda, et al., *J Clin Invest.* 100(7):1708 (1997); Allgood, et al., *Curr Opin Biotechnol.*, 8(4):474 (1997); Anders, et al., *J Biol Chem.*, 271(36):21758 (1996); Krishnan, et al., *Oncogene*, 13(1):125 (1996); Declercq, et al., *Cytokine*, 7(7):701 (1995); Bazzoni, et al., *Proc Natl Acad Sci U S A*, 92(12):5380 (1995); Ohashi, et al., *Proc Natl Acad Sci U S A*, 91(1):158 (1994); Desai, et al., *Cell*, 73(3):541 (1993); and Amara, et al., *Proc Natl Acad Sci U S A*, 94(20):10618 (1997)).

10
15
20
25
30

In addition to the extracellular domain and the cytosolic domain, these receptors have a transmembrane domain. As will be appreciated by those in the art, for chimeric death receptor genes, the transmembrane domain from any of the receptors can be used, although in general, it is preferred to use the transmembrane domain associated with the chosen cytosolic domain, to preserve the interaction of the transmembrane domain with other endogeneously signalling proteins.

Thus, preferred embodiments provide fusion nucleic acids that utilize the IL-4 inducible ϵ promoter linked to a death gene, particularly a chimeric death receptor gene, that requires a death ligand for cell killing.

20

Alternatively, inducible promoters can be linked to “one step” death genes, i.e. death genes that upon a certain threshold expression, will kill a cell without requiring a ligand or secondary signal. In this embodiment, the inducible promoter is preferably “leaky”, such that some small amount of death gene and a required secondary reporter gene such as a survival gene or a detection gene can be expressed. The cells that contain the death gene can then be selected on this basis, to avoid false positives. Once the presence of the construct is verified, candidate agents are added (and their presence preferably verified, using a detection or selection gene as well), and the promoter is induced. The population is then enriched for those cells that contain agents that inhibit the promoter, i.e. that will survive.

30

In a preferred embodiment, additional reporter genes are used, particularly when inducible death genes are used. In a preferred embodiment, the additional reporter gene is a selection gene. The cells containing the death gene and the drug selectable gene are grown; if the appropriate drug is added to the culture, only those cells containing the resistance gene (and 5 hence the death gene) survive. This ensures that the cells are expressing the death gene to decrease "false positives", i.e. cells that do not die because they do not contain the death gene.

In an additional preferred embodiment, the additional reporter gene is a labeling gene such as GFP. The use of a detection gene allows cells to be sorted to give a population enriched for those containing the construct. As outlined above, a preferred embodiment uses "leaky" 10 inducible promoters; that is, the cells are selected such that the IL-4 inducible promoter, even in the absence of IL-4 or IL-13, produces some GFP and death gene (for example, the Fas receptor constructs). In this embodiment, suitably "leaky" promoters are chosen such that 15 some GFP is expressed (preferably enough to select the cells expressing the construct from those that are not), but not enough death gene is produced to cause death. While preferred embodiments utilize death genes requiring the addition of a death ligand, it is well known that high levels of some death genes, even in the absence of death ligand, can cause death. Thus, 20 for example, high levels of Fas receptor expression can cause multimerization, and thus activation, even in the absence of the Fas ligand.

In a preferred embodiment, when two reporter genes are used, they are fused together in such a way as to only require a single promoter, and thus some way of functionally separating the two genes is preferred. This can be done on the RNA level or the protein level. Preferred 25 embodiments utilize either IRES sites (which allows the translation of two different genes on a single transcript (Kim, et al., "Construction of a Bifunctional mRNA in the Mouse By Using the Internal Ribosomal Entry Site of the Encephalomyocarditis Virus," *Molecular and Cellular Biology* 12(8):3636-3643 (Aug 1992) and McBratney, et al., "The Sequence Context of the Initiation Codon in the Encephalomyocarditis Virus Leader Modulates Efficiency of Internal Translation Initiation," *Current Opinion in Cell Biology* 5:961-965 (1993)), or a 30 protease cleavage site (which cleaves a protein translation product into two proteins).

Preferred protease cleavage sites include, but are not limited to, the 2a site (Ryan et al., J. Gen. Virol. 72:2727 (1991); Ryan et al., EMBO J. 13:928 (1994); Donnelly et al., J. Gen. Virol. 78:13 (1997); Hellen et al., *Biochem.* 28(26):9881 (1989); and Mattion et al., J. Virol. 70:8124 (1996), all of which are expressly incorporated by reference), prosequences of 5 retroviral proteases including human immunodeficiency virus protease and sequences recognized and cleaved by trypsin (EP 578472, Takasuga et al., J. Biochem. 112(5):652 (1992)) factor X_a (Gardella et al., J. Biol. Chem. 265(26):15854 (1990), WO 9006370), collagenase (J03280893, Tajima et al., J. Ferment. Bioeng. 72(5):362 (1991), WO 9006370), clostridain (EP 578472), subtilisin (including mutant H64A subtilisin, Forsberg et al., J. Protein Chem. 10(5):517 (1991), chymosin, yeast KEX2 protease (Bourbonnais et al., J. Bio. Chem. 263(30):15342 (1988), thrombin (Forsberg et al., *supra*; Abath et al., BioTechniques 10(2):178 (1991)), *Staphylococcus aureus* V8 protease or similar endoproteinase-Glu-C to 10 cleave after Glu residues (EP 578472, Ishizaki et al., Appl. Microbiol. Biotechnol. 36(4):483 (1992)), cleavage by NIa proteinase of tobacco etch virus (Parks et al., Anal. Biochem. 216(2):413 (1994)), endoproteinase-Lys-C (U.S. Patent No. 4,414,332) and 15 endoproteinase-Asp-N, *Neisseria* type 2 IgA protease (Pohlner et al., Bio/Technology 10(7):799-804 (1992)), soluble yeast endoproteinase yscF (EP 467839), chymotrypsin (Altman et al., Protein Eng. 4(5):593 (1991)), enteropeptidase (WO 9006370), lysostaphin, a polyglycine specific endoproteinase (EP 316748), and the like. See e.g. Marston, F.A.O. 20 (1986) Biol. Chem. J. **240**, 1-12.

In addition to the promoter of interest, such as an IL-4 inducible ϵ promoter and reporter 25 gene, the fusion nucleic acids may comprise additional components, including, but not limited to, other reporter genes, protein cleavage sites, internal ribosome entry (IRES) sites, AP-I sites, and other components as will be appreciated by those in the art.

In a preferred embodiment, foreign constructs comprising the IL-4 inducible ϵ promoter and the reporter gene are made. By "foreign" herein is meant that the fusion nucleic acids originates outside of the cells. That is, a recombinant nucleic acid is made that contains an 30 exogenous IL-4 inducible ϵ promoter and a reporter gene. Thus, in some circumstances, the

cells will contain both exogenous and endogenous IL-4 inducible ϵ promoters. By "recombinant nucleic acid" herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid by endonucleases, in a form not normally found in nature. Thus an isolated nucleic acid, in a linear form, a nucleic acid containing 5 components not normally joined, such as an IL-4 inducible promoter and a reporter gene, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are all considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will 10 replicate non-recombinantly, i.e. using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the 15 purposes of the invention. In this embodiment, any cells that express an IL-4 receptor that transduces the IL-4 signal to the nucleus and alters transcription can be used. Suitable cells include, but are not limited to, human cells and cell lines that show IL-4/13 inducible production of germline ϵ transcripts, including, but not limited to, DND39 (see Watanabe, 20 supra), MC-116, (Kumar, et al., "Human BCGF-12kD Functions as an Autocrine Growth Factor in Transformed B Cells," *Eur Cytokine Netw* 1(2):109 (1990)), CA-46 (Wang, et al., "UCN-01: A Potent Abrogator of G2 Checkpoint Function in Cancer Cells with Dirupted p53," *J Natl Cancer Inst* 88:956 (1996)).

This recombinant nucleic acid may introduced to a cell in a variety of ways, as will be appreciated by those in the art, including, but not limited to, CaPO_4 precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. The constructs may preferably stably integrate into the genome of the host cell (for example, with retroviral introduction, outlined 25 below), or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.).

In a preferred embodiment, the exogenous constructs, which may be in the form of an expression vector, are added as retroviral constructs, using techniques generally described in

PCT US97/01019 and PCT US97/01048, both of which are expressly incorporated by reference, and the examples.

In a preferred embodiment, the fusion construct comprises an endogeneous IL-4 inducible ϵ promoter and an exogeneous reporter gene; "endogeneous" in this context means originating within the cell. That is, gene "knock-in" constructions are made, whereby an exogeneous reporter gene as outlined herein is added, via homologous recombination, to the genome, such that the reporter gene is under the control of the endogeneous IL-4 inducible ϵ promoter.

This may be desirable to allow for the exploration and modulation of the full range of endogeneous regulation, i.e. regulatory elements (particularly those flanking the promoter) other than just the IL-4 inducible ϵ promoter fragment. Exemplary constructs are shown in Figures 5B and 5C, with GFP and BFP, although other reporter genes outlined herein may be used.

Homologous recombination may proceed in several ways. In one embodiment, traditional homologous recombination is done, with molecular biological techniques such as PCR being done to find the correct insertions. For example, gene "knock-ins" may be done as is known in the art, for example see Westphal et al., Current Biology 7:R530-R533 (1997), and references cited therein, all of which are expressly incorporated by reference. The use of recA mediated systems may also be done, see PCT US93/03868, hereby expressly incorporated by reference.

Alternatively, and preferably, the selection of the "knock ins" are done by FACS on the basis of the incorporation of a reporter gene. Thus, in a preferred embodiment, a first homologous recombination event is done to put a first reporter gene, such as GFP, into at least one allele of the cell genome. Preferably, this is a cell type that exhibits IL-4 inducible production of at least germline ϵ transcripts, so that the cells may be tested by IL-4 production for reporter gene expression. Suitable cells include, but are not limited to, human cells and cell lines that show IL-4/13 inducible production of germline ϵ transcripts, including, but not limited to, DND39 (see Watanabe, *supra*), MC-116, (Kumar, et al., "Human BCGF-12kD Functions as

an autocrine Growth Factor in Transformed B Cells," *Eur Cytokine Netw* 1(2):109 (1990)), CA-46 (Wang, et al., "UCN-01: A Potent Abrogator of G2 Checkpoint Function in Cancer Cells with Dirupted p53," *J Natl Cancer Inst* 88:956 (1996)). As is noted herein, the ability of MC-116 and CA-46 cells to produce germline ϵ transcripts upon IL-4/13 induction was 5 not known prior to the present invention. Thus, preferred embodiments provide MC-116 and/or CA-46 cells comprising recombinant nucleic acid reporter constructs are outlined herein.

10 In a preferred embodiment, once a first endogenous promoter has been combined with an exogenous reporter construct, a second homologous recombination event may be done, preferably using a second reporter gene different from the first, such as BFP, to target the other allele of the cell genome, and tested as above.

15 Generally, IL-4 induction of the reporter genes will indicate the correct placement of the genes, which can be confirmed via sequencing such as PCR sequencing or Southern blot hybridization. In addition, preferred embodiments utilize prescreening steps to remove "leaky" cells, i.e. those showing constitutive expression of the reporter gene.

20 Thus, in a preferred embodiment, the invention provides cell lines that contain fusion nucleic acids comprising IL-4 inducible ϵ promoter operably connected to at least one reporter gene. Once made, the cell lines comprising these reporter constructs are used to screen candidate bioactive agents for the ability to modulate the production of IgE, as is outlined below.

25 The term "candidate bioactive agent" or "exogeneous compound" as used herein describes any molecule, *e.g.*, protein, oligopeptide, small organic molecule, polysaccharide, polynucleotide. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.*, at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at 5 least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polycyclic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including 10 peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof. Particularly preferred are peptides.

15 Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means. Known pharmacological agents 20 may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification to produce structural analogs.

25 In a preferred embodiment, the candidate bioactive agents are proteins. By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline and noreleucine are considered amino acids for the purposes of the invention. "Amino acid" also includes imino acid residues such as proline and hydroxyproline. The side chains may be in either the (R) or the (S) configuration. In the 30 preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally

occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard *in vivo* degradations.

5 In a preferred embodiment, the candidate bioactive agents are naturally occurring proteins or fragments of naturally occurring proteins. Thus, for example, cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of prokaryotic and eucaryotic proteins may be made for screening in the systems described herein. Particularly preferred in this embodiment are libraries of bacterial, fungal, viral, and mammalian proteins, with the latter being preferred, and human proteins being especially preferred.

10 In a preferred embodiment, the candidate bioactive agents are peptides of from about 5 to about 30 amino acids, with from about 5 to about 20 amino acids being preferred, and from about 7 to about 15 being particularly preferred. The peptides may be digests of naturally occurring proteins as is outlined above, random peptides, or "biased" random peptides. By "randomized" or grammatical equivalents herein is meant that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate any nucleotide or amino acid at any position. The synthetic process can be 15 designed to generate randomized proteins or nucleic acids, to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized candidate bioactive proteinaceous agents.

20 In one embodiment, the library is fully randomized, with no sequence preferences or constants at any position. In a preferred embodiment, the library is biased. That is, some 25 positions within the sequence are either held constant, or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or amino acid residues are randomized within a defined class, for example, of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of

cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., or to purines, etc.

In a preferred embodiment, the candidate bioactive agents are nucleic acids. By "nucleic

5 acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramidate

10 (Beaucage, *et al.*, Tetrahedron, **49**(10):1925 (1993) and references therein; Letsinger, J. Org. Chem., **35**:3800 (1970); Sprinzl, *et al.*, Eur. J. Biochem., **81**:579 (1977); Letsinger, *et al.*, Nucl. Acids Res., **14**:3487 (1986); Sawai, *et al.*, Chem. Lett., 805 (1984), Letsinger, *et al.*, J. Am. Chem. Soc., **110**:4470 (1988); and Pauwels, *et al.*, Chemica Scripta, **26**:141 (1986)), phosphorothioate (Mag, *et al.*, Nucleic Acids Res., **19**:1437 (1991); and U.S. Patent No.

15 5,644,048), phosphorodithioate (Brui, *et al.*, J. Am. Chem. Soc., **111**:2321 (1989)), O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc., **114**:1895 (1992); Meier, *et al.*, Chem. Int. Ed. Engl., **31**:1008 (1992); Nielsen, Nature, **365**:566 (1993); Carlsson, *et al.*, Nature, **380**:207 (1996), all of which are incorporated by reference)). Other analog nucleic acids include those with positive

20 backbones (Denpny, *et al.*, Proc. Natl. Acad. Sci. USA, **92**:6097 (1995)); non-ionic backbones (U.S. Patent Nos. 5,386,023; 5,637,684; 5,602,240; 5,216,141; and 4,469,863; Kiedrowski, *et al.*, Angew. Chem. Intl. Ed. English, **30**:423 (1991); Letsinger, *et al.*, J. Am. Chem. Soc., **110**:4470 (1988); Letsinger, *et al.*, Nucleoside & Nucleotide, **13**:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook; Mesmaeker, *et al.*, Bioorganic & Medicinal Chem. Lett., **4**:395 (1994); Jeffs, *et al.*, J. Biomolecular NMR, **34**:17 (1994); Tetrahedron Lett., **37**:743 (1996)) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook.

30 Nucleic acids containing one or more carbocyclic sugars are also included within the

definition of nucleic acids (see Jenkins, *et al.*, *Chem. Soc. Rev.*, (1995) pp. 169-176). Several nucleic acid analogs are described in Rawls, C & E News, June 2, 1997, page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of additional moieties such as

5 labels, or to increase the stability and half-life of such molecules in physiological environments. In addition, mixtures of naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthanine hypoxanthanine, isocytosine, isoguanine, etc.

10

15.0 As described above generally for proteins, nucleic acid candidate bioactive agents may be naturally occurring nucleic acids, random nucleic acids, or "biased" random nucleic acids. For example, digests of prokaryotic or eucaryotic genomes may be used as is outlined above for proteins.

20 In a preferred embodiment, the candidate bioactive agents are organic chemical moieties, a wide variety of which are available in the literature.

In a preferred embodiment, a library of different candidate bioactive agents are used. Preferably, the library should provide a sufficiently structurally diverse population of

25 randomized agents to effect a probabilistically sufficient range of diversity to allow binding to a particular target. Accordingly, an interaction library should be large enough so that at least one of its members will have a structure that gives it affinity for the target. Although it is difficult to gauge the required absolute size of an interaction library, nature provides a hint with the immune response: a diversity of 10^7 - 10^8 different antibodies provides at least one

30 combination with sufficient affinity to interact with most potential antigens faced by an

organism. Published *in vitro* selection techniques have also shown that a library size of 10^7 to 10 8 is sufficient to find structures with affinity for the target. A library of all combinations of a peptide 7 to 20 amino acids in length, such as generally proposed herein, has the potential to code for 20^7 (10^9) to 20^{20} . Thus, with libraries of 10^7 to 10 8 different molecules the present 5 methods allow a "working" subset of a theoretically complete interaction library for 7 amino acids, and a subset of shapes for the 20^{20} library. Thus, in a preferred embodiment, at least 10 6 , preferably at least 10 7 , more preferably at least 10 8 and most preferably at least 10 9 different sequences are simultaneously analyzed in the subject methods. Preferred methods maximize library size and diversity.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

The candidate bioactive agents are combined or added to a cell or population of cells. Suitable cell types for different embodiments are outlined above. By "population of cells" herein is meant at least two cells, with at least about 10 5 being preferred, at least about 10 6 being particularly preferred, and at least about 10 7 , 10 8 and 10 9 being especially preferred.

The candidate bioactive agent and the cells are combined. As will be appreciated by those in the art, this may be accomplished in any number of ways, including adding the candidate agents to the surface of the cells, to the media containing the cells, or to a surface on which the cells are growing or in contact with; adding the agents into the cells, for example by using vectors 20 that will introduce the agents into the cells (i.e. when the agents are nucleic acids or proteins).

In a preferred embodiment, the candidate bioactive agents are either nucleic acids or proteins (proteins in this context includes proteins, oligopeptides, and peptides) that are introduced into the host cells using retroviral vectors, as is generally outlined in PCT US97/01019 and 25 PCT US97/01048, both of which are expressly incorporated by reference. Generally, a library of retroviral vectors is made using retroviral packaging cell lines that are helper-defective and are capable of producing all the necessary trans proteins, including gag, pol and env, and RNA molecules that have in *cis* the Ψ packaging signal. Briefly, the library is generated in a retrovirus DNA construct backbone; standard oligonucleotide synthesis is done to generate 30 either the candidate agent or nucleic acid encoding a protein, for example a random peptide,

using techniques well known in the art. After generation of the DNA library, the library is cloned into a first primer. The first primer serves as a "cassette", which is inserted into the retroviral construct. The first primer generally contains a number of elements, including for example, the required regulatory sequences (e.g. translation, transcription, promoters, etc), fusion partners, restriction endonuclease (cloning and subcloning) sites, stop codons (preferably in all three frames), regions of complementarity for second strand priming (preferably at the end of the stop codon region as minor deletions or insertions may occur in the random region), etc.

10 A second primer is then added, which generally consists of some or all of the complementarity region to prime the first primer and optional necessary sequences for a second unique restriction site for subcloning. DNA polymerase is added to make double-stranded oligonucleotides. The double-stranded oligonucleotides are cleaved with the appropriate subcloning restriction endonucleases and subcloned into the target retroviral vectors, described below.

150

Any number of suitable retroviral vectors may be used. Generally, the retroviral vectors may include: selectable marker genes under the control of internal ribosome entry sites (IRES) that greatly facilitates the selection of cells expressing peptides at uniformly high levels; and 20 promoters driving expression of a second gene, placed in sense or anti-sense relative to the 5' LTR. Suitable selection genes include, but are not limited to, neomycin, blastocidin, bleomycin, puromycin, and hygromycin resistance genes, as well as self-fluorescent markers such as green fluorescent protein, enzymatic markers such as lacZ, and surface proteins such as CD8, etc.

Preferred vectors include a vector based on the murine stem cell virus (MSCV) (see Hawley et al., *Gene Therapy* 1:136 (1994)) and a modified MFG virus (Rivere et al., *Genetics* 92:6733 (1995)), and pBABE, outlined in the examples.

The retroviruses may include inducible and constitutive promoters for the expression of the candidate agent (to be distinguished from the IL-4 inducible ϵ promoter). For example, there are situations wherein it is necessary to induce peptide expression only during certain phases of the selection process. A large number of both inducible and constitutive promoters are

5 known.

In addition, it is possible to configure a retroviral vector to allow inducible expression of retroviral inserts after integration of a single vector in target cells; importantly, the entire system is contained within the single retrovirus. Tet-inducible retroviruses have been 10 designed incorporating the Self-Inactivating (SIN) feature of 3' LTR enhancer/promoter retroviral deletion mutant (Hoffman et al., PNAS USA 93:5185 (1996)). Expression of this vector in cells is virtually undetectable in the presence of tetracycline or other active analogs. However, in the absence of Tet, expression is turned on to maximum within 48 hours after induction, with uniform increased expression of the whole population of cells that harbor the 15 inducible retrovirus, indicating that expression is regulated uniformly within the infected cell population. A similar, related system uses a mutated Tet DNA-binding domain such that it bound DNA in the presence of Tet, and was removed in the absence of Tet. Either of these systems is suitable.

20 In a preferred embodiment, the candidate bioactive agents are linked to a fusion partner. By “fusion partner” or “functional group” herein is meant a sequence that is associated with the candidate bioactive agent, that confers upon all members of the library in that class a common function or ability. Fusion partners can be heterologous (i.e. not native to the host cell), or synthetic (not native to any cell). Suitable fusion partners include, but are not limited to: a) 25 presentation structures, as defined below, which provide the candidate bioactive agents in a conformationally restricted or stable form; b) targeting sequences, defined below, which allow the localization of the candidate bioactive agent into a subcellular or extracellular compartment, particularly a nuclear localization sequence (NLS); c) rescue sequences as defined below, which allow the purification or isolation of either the candidate bioactive 30 agents or the nucleic acids encoding them; d) stability sequences, which confer stability or

protection from degradation to the candidate bioactive agent or the nucleic acid encoding it, for example resistance to proteolytic degradation; e) dimerization sequences, to allow for peptide dimerization; f) reporter genes (preferably a labeling gene or a survival gene); or g) any combination of a), b), c), d), e), or f) as well as linker sequences as needed.

5

In a preferred embodiment, the fusion partner is a presentation structure. By "presentation structure" or grammatical equivalents herein is meant a sequence, which, when fused to candidate bioactive agents, causes the candidate agents to assume a conformationally restricted form. Proteins interact with each other largely through conformationally constrained domains. Although small peptides with freely rotating amino and carboxyl termini can have potent functions as is known in the art, the conversion of such peptide structures into pharmacologic agents is difficult due to the inability to predict side-chain positions for peptidomimetic synthesis. Therefore the presentation of peptides in conformationally constrained structures will benefit both the later generation of pharmaceuticals and will also likely lead to higher affinity interactions of the peptide with the target protein. This fact has been recognized in the combinatorial library generation systems using biologically generated short peptides in bacterial phage systems. A number of workers have constructed small domain molecules in which one might present randomized peptide structures.

20

While the candidate bioactive agents may be either nucleic acid or peptides, presentation structures are preferably used with peptide candidate agents. Thus, synthetic presentation structures, i.e. artificial polypeptides, are capable of presenting a randomized peptide as a conformationally-restricted domain. Generally such presentation structures comprise a first portion joined to the N-terminal end of the randomized peptide, and a second portion joined to the C-terminal end of the peptide; that is, the peptide is inserted into the presentation structure, although variations may be made, as outlined below. To increase the functional isolation of the randomized expression product, the presentation structures are selected or designed to have minimal biologically activity when expressed in the target cell.

30

Preferred presentation structures maximize accessibility to the peptide by presenting it on an exterior loop. Accordingly, suitable presentation structures include, but are not limited to, minibody structures, loops on beta-sheet turns and coiled-coil stem structures in which residues not critical to structure are randomized, zinc-finger domains, cysteine-linked (disulfide) structures, transglutaminase linked structures, cyclic peptides, B-loop structures, helical barrels or bundles, leucine zipper motifs, etc.

5

In a preferred embodiment, the presentation structure is a coiled-coil structure, allowing the presentation of the randomized peptide on an exterior loop. See, for example, Myszka et al., Biochem. 33:2362-2373 (1994), hereby incorporated by reference). Using this system investigators have isolated peptides capable of high affinity interaction with the appropriate target. In general, coiled-coil structures allow for between 6 to 20 randomized positions.

10

A preferred coiled-coil presentation structure is as follows:

15

MGCAALESEVSALESEVAALGRGDMPLAAVKSKLSAVKSKLASVKSKLAACGPP. The underlined regions represent a coiled-coil leucine zipper region defined previously (see Martin et al., EMBO J. 13(22):5303-5309 (1994), incorporated by reference).

20

The bolded GRGDMP region represents the loop structure and when appropriately replaced with randomized peptides (i.e. candidate bioactive agents, generally depicted herein as (X)_n), where X is an amino acid residue and n is an integer of at least 5 or 6) can be of variable length. The replacement of the bolded region is facilitated by encoding restriction endonuclease sites in the underlined regions, which allows the direct incorporation of randomized oligonucleotides at these positions. For example, a preferred embodiment generates a XhoI site at the double underlined LE site and a HindIII site at the double-underlined KL site.

25

30

In a preferred embodiment, the presentation structure is a minibody structure. A "minibody" is essentially composed of a minimal antibody complementarity region. The minibody presentation structure generally provides two randomizing regions that in the folded protein are presented along a single face of the tertiary structure. See for example Bianchi et al., J.

Mol. Biol. 236(2):649-59 (1994), and references cited therein, all of which are incorporated by reference). Investigators have shown this minimal domain is stable in solution and have used phage selection systems in combinatorial libraries to select minibodies with peptide regions exhibiting high affinity, $K_d = 10^{-7}$, for the pro-inflammatory cytokine IL-6.

5

A preferred minibody presentation structure is as follows:

MGRNSQATSGFTSHEYMEWVRGGEYIAASRHKHNKYTTEYSASVKGRYIVSRDT
SQSILYLQKKKGPP. The bold, underline regions are the regions which may be randomized.

10 The italicized phenylalanine must be invariant in the first randomizing region. The entire peptide is cloned in a three-oligonucleotide variation of the coiled-coil embodiment, thus allowing two different randomizing regions to be incorporated simultaneously. This embodiment utilizes non-palindromic BstXI sites on the termini.

15 In a preferred embodiment, the presentation structure is a sequence that contains generally two cysteine residues, such that a disulfide bond may be formed, resulting in a conformationally constrained sequence. This embodiment is particularly preferred when secretory targeting sequences are used. As will be appreciated by those in the art, any number of random sequences, with or without spacer or linking sequences, may be flanked with cysteine residues. In other embodiments, effective presentation structures may be generated 20 by the random regions themselves. For example, the random regions may be "doped" with cysteine residues which, under the appropriate redox conditions, may result in highly crosslinked structured conformations, similar to a presentation structure. Similarly, the randomization regions may be controlled to contain a certain number of residues to confer β -sheet or α -helical structures.

25

In a preferred embodiment, the fusion partner is a targeting sequence that targets the candidate bioactive agent to a particular subcellular location. As will be appreciated by those in the art, the localization of proteins within a cell is a simple method for increasing effective concentration and determining function. The concentration of a protein can also be simply 30 increased by nature of the localization. Shuttling the proteins into the nucleus confines them

to a smaller space thereby increasing concentration. While other targeting sequences such as targeting sequences to the Golgi, endoplasmic reticulum, nuclear membrane, mitochondria, secretory vesicles, lysosome, and cellular membrane may be used, a preferred embodiment uses targeting sequences to the nucleus, i.e. a nuclear localization signal (NLS).

5

10
15
20
25
30

In a preferred embodiment, the targeting sequence is a nuclear localization signal (NLS). NLSs are generally short, positively charged (basic) domains that serve to direct the entire protein in which they occur to the cell's nucleus. Numerous NLS amino acid sequences have been reported including single basic NLS's such as that of the SV40 (monkey virus) large T Antigen (Pro Lys Lys Lys Arg Lys Val), Kalderon (1984), et al., Cell, 39:499-509; the human retinoic acid receptor- β nuclear localization signal (ARRRRP); NF κ B p50 (EEVQRKRQKL; Ghosh et al., Cell 62:1019 (1990); NF κ B p65 (EEKRKRTYE; Nolan et al., Cell 64:961 (1991); and others (see for example Boulikas, J. Cell. Biochem. 55(1):32-58 (1994), hereby incorporated by reference) and double basic NLS's exemplified by that of the *Xenopus* (African clawed toad) protein, nucleoplasmin (Ala Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Leu Asp), Dingwall, et al., Cell, 30:449-458, 1982 and Dingwall, et al., J. Cell Biol., 107:641-849; 1988). Numerous localization studies have demonstrated that NLSs incorporated in synthetic peptides or grafted onto reporter proteins not normally targeted to the cell nucleus cause these peptides and reporter proteins to be concentrated in the nucleus. See, for example, Dingwall, and Laskey, Ann, Rev. Cell Biol., 2:367-390, 1986; Bonnerot, et al., Proc. Natl. Acad. Sci. USA, 84:6795-6799, 1987; Galileo, et al., Proc. Natl. Acad. Sci. USA, 87:458-462, 1990.

20

25

30

In a preferred embodiment, the fusion partner is a rescue sequence. A rescue sequence is a sequence which may be used to purify or isolate either the candidate agent or the nucleic acid encoding it. Thus, for example, peptide rescue sequences include purification sequences such as the His₆ tag for use with Ni affinity columns and epitope tags for detection, immunoprecipitation or FACS (fluorescence-activated cell sorting). Suitable epitope tags include myc (for use with the commercially available 9E10 antibody), the BSP biotinylation target sequence of the bacterial enzyme BirA, flu tags, lacZ, and GST.

Alternatively, the rescue sequence may be a unique oligonucleotide sequence which serves as a probe target site to allow the quick and easy isolation of the retroviral construct, via PCR, related techniques, or hybridization.

5 In a preferred embodiment, the fusion partner is a stability sequence to confer stability to the candidate bioactive agent or the nucleic acid encoding it. Thus, for example, peptides may be stabilized by the incorporation of glycines after the initiation methionine (MG or MGG0), for protection of the peptide to ubiquitination as per Varshavsky's N-End Rule, thus conferring long half-life in the cytoplasm. Similarly, two prolines at the C-terminus impart peptides that are largely resistant to carboxypeptidase action. The presence of two glycines prior to the prolines impart both flexibility and prevent structure initiating events in the di-proline to be propagated into the candidate peptide structure. Thus, preferred stability sequences are as follows: MG(X)_nGGPP, where X is any amino acid and n is an integer of at least four.

15 In one embodiment, the fusion partner is a dimerization sequence. A dimerization sequence allows the non-covalent association of one random peptide to another random peptide, with sufficient affinity to remain associated under normal physiological conditions. This effectively allows small libraries of random peptides (for example, 10^4) to become large libraries if two peptides per cell are generated which then dimerize, to form an effective 20 library of 10^8 ($10^4 \times 10^4$). It also allows the formation of longer random peptides, if needed, or more structurally complex random peptide molecules. The dimers may be homo- or heterodimers.

25 Dimerization sequences may be a single sequence that self-aggregates, or two sequences, each of which is generated in a different retroviral construct. That is, nucleic acids encoding both a first random peptide with dimerization sequence 1, and a second random peptide with dimerization sequence 2, such that upon introduction into a cell and expression of the nucleic acid, dimerization sequence 1 associates with dimerization sequence 2 to form a new random peptide structure.

Suitable dimerization sequences will encompass a wide variety of sequences. Any number of protein-protein interaction sites are known. In addition, dimerization sequences may also be elucidated using standard methods such as the yeast two hybrid system, traditional biochemical affinity binding studies, or even using the present methods.

5

10
15
20
25
30

In a preferred embodiment, the fusion partner is a detection gene, preferably a labeling gene or a survival gene. That is, it is desirable to know that the candidate bioactive agent is a) present and b) being expressed. Thus, preferred embodiments utilize fusion constructs utilizing genes that allow the detection of cells that contain candidate bioactive agents, as is generally outlined in the Examples, and shown in Figure 10. Preferred detection genes include, but are not limited to, GFP, BFP, YFP, RFP, luciferase, and β -galactosidase. Preferred embodiments utilize detection genes that are different from the reporter genes used to determine whether the IL-4 inducible promoter is inhibited; that is, if a GFP reporter gene is used, preferably a non-GFP detection gene is used. This allows cell enrichment using FACS that can distinguish between cells containing candidate agents and those that do not, as well distinguishing cells containing candidate agents that do not inhibit the promoter and cells containing candidate agents that do inhibit the promoter.

In a preferred embodiment, as for the other constructs outlined herein, when a detection gene fusion partner is used with nucleic acid encoding a peptide candidate agent (which may also include other fusion partners as described herein), the two nucleic acids are fused together in such a way as to only require a single promoter, i.e. using either an IRES site or a protease cleavage site such as 2a. A preferred embodiment is depicted in Figure 10B.

The fusion partners may be placed anywhere (i.e. N-terminal, C-terminal, internal) in the structure as the biology and activity permits.

In a preferred embodiment, the fusion partner includes a linker or tethering sequence, as generally described in PCT US 97/01019, that can allow the candidate agents to interact with potential targets unhindered. For example, when the candidate bioactive agent is a peptide,

useful linkers include glycine-serine polymers (including, for example, (GS)_n, (GSGGS)_n and (GGGS)_n, where n is an integer of at least one), glycine-alanine polymers, alanine-serine polymers, and other flexible linkers such as the tether for the shaker potassium channel, and a large variety of other flexible linkers, as will be appreciated by those in the art. Glycine-
5 serine polymers are preferred since both of these amino acids are relatively unstructured, and therefore may be able to serve as a neutral tether between components. Secondly, serine is hydrophilic and therefore able to solubilize what could be a globular glycine chain. Third, similar chains have been shown to be effective in joining subunits of recombinant proteins such as single chain antibodies.

10
15
20
25
30

In addition, the fusion partners, including presentation structures, may be modified, randomized, and/or matured to alter the presentation orientation of the randomized expression product. For example, determinants at the base of the loop may be modified to slightly modify the internal loop peptide tertiary structure, which maintaining the randomized amino acid sequence.

In a preferred embodiment, combinations of fusion partners are used. Thus, for example, any number of combinations of presentation structures, targeting sequences, rescue sequences, and stability sequences may be used, with or without linker sequences.

20 Thus, candidate agents can include these components, and may then be used to generate a library of fragments, each containing a different random nucleotide sequence that may encode a different peptide. The ligation products are then transformed into bacteria, such as *E. coli*, and DNA is prepared from the resulting library, as is generally outlined in Kitamura, PNAS
25 USA 92:9146-9150 (1995), hereby expressly incorporated by reference.

Delivery of the library DNA into a retroviral packaging system results in conversion to infectious virus. Suitable retroviral packaging system cell lines include, but are not limited to, the Bing and BOSC23 cell lines described in WO 94/19478; Soneoka et al., Nucleic Acid
30 Res. 23(4):628 (1995); Finer et al., Blood 83:43 (1994); Pheonix packaging lines such as

PhiNX-eco and PhiNX-ampho, described below; 292T + gag-pol and retrovirus envelope; PA317; and cell lines outlined in Markowitz et al., Virology 167:400 (1988), Markowitz et al., J. Virol. 62:1120 (1988), Li et al., PNAS USA 93:11658 (1996), Kinsella et al., Human Gene Therapy 7:1405 (1996), all of which are incorporated by reference. Preferred systems 5 include PhiNX-eco and PhiNX-ampho or similar cell lines, disclosed in PCT US97/01019.

In general, the candidate agents are added to the cells under reaction conditions that favor agent-target interactions. Generally, this will be physiological conditions. Incubations may be performed at any temperature which facilitates optimal activity, typically between 4 and 10 40°C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high throughput screening. Typically between 0.1 and 1 hour will be sufficient. Excess reagent is generally removed or washed away.

15 A variety of other reagents may be included in the assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite binding.

20 Once the candidate agents have been introduced or combined with the cells containing the fusion constructs, the IL-4 inducible ϵ promoter is induced. Alternatively, the promoter is induced prior to the addition of the candidate bioactive agents, or simultaneously. This is generally done as is known in the art, and involves the addition of IL-4 or IL-13 to the cells at 25 a concentration of not less than 5 units/ml with 200 units/ml being most preferred. Addition of IL-4 or IL-13 is usually 24-48 hours after the bioactive agents are added.

30 The presence or absence of the reporter gene is then detected. This may be done in a number of ways, as will be appreciated by those in the art, and will depend in part on the reporter gene. For example, cells expressing a label reporter gene, such as GFP, can be distinguished

from those not expressing the gene, and preferably sorted (enriched by FACS) on this basis. Similarly, cells expressing the death gene will die, leaving only cells that have inhibited promotion of the expression of the gene, etc. In general, the cells that express the reporter gene (i.e. non-inhibited IL-4 inducible ϵ promoter) and separated from those that do not (i.e. 5 the IL-4 inducible ϵ promoter was inhibited). This may be done using FACS, lysis selection using complements, cell cloning, scanning by a Fluorimager, growth under drug resistance, enhanced growth, etc.

10 In a preferred embodiment, for example when the reporter gene is a death gene, sorting of cells containing bioactive agents that inhibit the IL-4 inducible ϵ promoter (and thus do not turn on the death gene) from those cells that contain candidate agents that do not inhibit the promoter is simple: only those surviving cells contain such an agent.

15 In a preferred embodiment, the presence or absence of the reporter gene is determined using a fluorescent-activated cell sorter (FACS). In general, the expression of the reporter gene comprising a label (or allowing the use of a label) is optimized to allow for efficient enrichment by FACS. Thus, for example, in general, 10 to 1000 fluores per sorting event are needed; i.e. per cell, with from about 100 to 1000 being preferred, and from 500 to 1000 being especially preferred. This can be accomplished by amplifying the signal per reporter 20 gene, i.e. have each second label comprise multiple fluores, or by having a high density of reporter genes per cell; or a combination of both.

25 In a preferred embodiment, the cells are sorted at very high speeds, for example greater than about 5,000 sorting events per sec, with greater than about 10,000 sorting events per sec being preferred, and greater than about 25,000 sorting events per second being particularly preferred, with speeds of greater than about 50,000 to 100,000 being especially preferred. The use of multiple laser paths allows sort accuracy of 1 in 10^6 with better than 70% accuracy.

The sorting results in a population of cells containing the reporter protein (i.e. the promoter was not inhibited) and at least one population of cells without the reporter protein (i.e. the promoter was inhibited). The absence of the reporter protein is indicative that at least one candidate bioactive agent is a bioactive agent that inhibits the IL-4 inducible ϵ promoter.

5

10
15
20
25

In addition to screening methods utilizing the reporter constructs described above, the invention also provides methods for screening candidate agents for the ability to modulate IgE production. By "modulating IgE production" herein is meant either an increase or a decrease in IgE production, as quantified by the amount of IgE protein made. In this embodiment, cells that have already switched to the ϵ heavy chain region can no longer be blocked at the earlier phase of IgE production. This is especially important for memory B cells that maintain their capacity to secrete IgE and are long lived. Thus, in this embodiment, candidate agents are screened to identify compounds that can block IgE at the level of ϵ heavy chain transcription, translation, assembly and trafficking, to prevent the terminal stages of IgE production. In this embodiment, a candidate bioactive agent is combined with a cell capable of expressing IgE, preferably surface IgE. Preferred cells include, but are not limited to, cells that produce surface IgE such as the U266 cell line (Lagging, et al., "Distribution of Plasma Cell Markers and Intracellular IgE in Cell Line U266," *Immunology Letters* 49:71 (1996)).

20

25

The candidate agent and the cells are combined, as outlined above, and the cells screened for alterations in the amount of IgE produced, as compared to the amount produced in the absence of the candidate bioactive agent. This may be done using standard IgE labeling techniques, including, but not limited to, the use of anti-IgE antibodies, that may be either directly or indirectly labeled, for example through the use of fluorescent anti-IgE antibodies or fluorescent secondary antibodies, and through the use of IgE fusion proteins, as outlined below.

30

In a preferred embodiment, the amount of IgE produced is determined through the use of IgE fusion proteins; that is, the IgE is produced as a fusion protein comprising the IgE protein,

specifically at least the ϵ heavy chain, and a detectable protein such as is generally outlined above for label reporter genes. In a preferred embodiment, gene "knock in" cell lines are produced, as outlined above and shown in the Figures. In this embodiment, a first label gene, such as the gene for green fluorescent protein (GFP), is fused to the secretory exon of IgE to 5 label secretory IgE heavy chains green. In a preferred embodiment, a second label gene, such as the gene for blue fluorescent protein (BFP), is attached to the M2 exon to label membrane IgE heavy chains blue. This is preferred as it allows discrimination between mRNA processing and translation of secretory versus membrane ϵ -heavy chain transcripts. Suitable label genes for this embodiment include, but are not limited to, GFP, BFP, YFP and RFP.

10

Accordingly, the present invention provides cell lines that produce fusion proteins comprising IgE (either secreted or membrane bound) fused to a label protein, preferably a fluorescent protein.

150

10

5

10

15

20

25

30

In yet another preferred embodiment, the invention provides methods of identifying proteins that bind to all or part of the switch ϵ region (Figure 2B). The general idea is to use a "one hybrid" system to identify proteins that bind to all or part of the switch ϵ region. To this end, the present invention provides compositions comprising a test vector and a reporter vector, and cells containing these vectors. These cells may be yeast, such as YM4271 or any yeast cell lines that reporter constructs can be inserted into.

By "vector" or "episome" herein is meant a replicon used for the transformation of host cells. The vectors may be either self-replicating extrachromosomal vectors ("plasmids") or vectors which integrate into a host genome. A preferred embodiment utilizes retroviral vectors, as is more fully described below.

Suitable vectors will depend on the host cells used. For use of the system in yeast, suitable vectors are known in the art and include, but are not limited to, pHisi-1 and pLacZi (Clontech Cat #K1603-1) (Li, et al., "Isolation of ORC6, A Component of the Yeast Origin of Recognition Complex By a One-Hybrid System," *Science* 262:1870-1873 (1993); Liu, et

al. "Identifying DNA-Binding Sites and Analyzing DNA-Binding Domains Using a Yeast Selection System," In: *Methods: A Companion to Methods in Enzymology* 5:125-137 (1993), Luo, et al., "Cloning and Analysis of DNA-Binding Proteins By Yeast One-Hybrid and One-Two-Hybrid Systems," *Biotechniques* 20:564-568 (1996), and Strubin, et al., "OBF-1, A 5 Novel B Cell-Specific Coactivator That Stimulates Immunoglobulin Promoter Activity Through Association with Octamer-Binding Proteins," *Cell* 80:497-506 (1995)). Yeast expression systems are well known in the art, and include expression vectors for *Saccharomyces cerevisiae*, *Candida albicans* and *C. maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis* and *K. lactis*, *Pichia guillermondi* and *P. pastoris*, 10 *Schizosaccharomyces pombe*, and *Yarrowia lipolytica*. Preferred promoter sequences for expression in yeast include the inducible GAL1,10 promoter, the promoters from alcohol dehydrogenase, enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase, hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, pyruvate kinase, and the acid phosphatase gene. Yeast selectable markers include ADE2, 15 HIS4, LEU2, TRP1, and ALG7, which confers resistance to tunicamycin; the neomycin phosphotransferase gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the presence of copper ions.

20 For non-retroviral mammalian cell embodiments, suitable vectors are derived from any number of known vectors, including, but not limited to, pCEP4 (Invitrogen), pCI-NEO (Promega), and pBI-EGFP (Clontech). Basically, any mammalian expression vectors with strong promoters such as CMV can be used to construct test vectors.

25 In a preferred embodiment, one or more retroviral vectors are used. Currently, the most efficient gene transfer methodologies harness the capacity of engineered viruses, such as retroviruses, to bypass natural cellular barriers to exogenous nucleic acid uptake. The use of recombinant retroviruses was pioneered by Richard Mulligan and David Baltimore with the Psi-2 lines and analogous retrovirus packaging systems, based on NIH 3T3 cells (see Mann et al., *Cell* 33:153-159 (1993), hereby incorporated by reference). Such helper-defective 30 packaging lines are capable of producing all the necessary trans proteins -gag, pol, and env-

that are required for packaging, processing, reverse transcription, and integration of recombinant genomes. Those RNA molecules that have in cis the Ψ packaging signal are packaged into maturing virions.

5 Retroviruses are preferred for a number of reasons. First, their derivation is easy. Second, unlike Adenovirus-mediated gene delivery, expression from retroviruses is long-term (adenoviruses do not integrate). Adeno-associated viruses have limited space for genes and regulatory units and there is some controversy as to their ability to integrate. Retroviruses therefore offer the best current compromise in terms of long-term expression, genomic 10 flexibility, and stable integration, among other features. The main advantage of retroviruses is that their integration into the host genome allows for their stable transmission through cell division. This ensures that in cell types which undergo multiple independent maturation steps, such as hematopoietic cell progression, the retrovirus construct will remain resident and continue to express. In addition, transfection efficiencies can be extremely high, thus 15 obviating the need for selection genes in some cases.

A particularly well suited retroviral transfection system is described in Mann et al., *supra*:
Pear et al., PNAS USA 90(18):8392-6 (1993); Kitamura et al., PNAS USA 92:9146-9150
(1995); Kinsella et al., Human Gene Therapy 7:1405-1413; Hofmann et al., PNAS USA
20 93:5185-5190; Choate et al., Human Gene Therapy 7:2247 (1996); WO 94/19478; PCT
US97/01019, and references cited therein, all of which are incorporated by reference.

Any number of suitable retroviral vectors may be used. Preferred retroviral vectors include a 25 vector based on the murine stem cell virus (MSCV) (see Hawley et al., Gene Therapy 1:136 (1994)) and a modified MFG virus (Rivere et al., Genetics 92:6733 (1995)), and pBABE (see PCT US97/01019, incorporated by reference). Particularly preferred vectors are shown in Figure 11.

As for the other vectors, the retroviral vectors may include inducible and constitutive 30 promoters. Constitutive promoters are preferred for the bait and test vectors, and include, but

are not limited to, CMV, SV40, $\text{Sr}\alpha$, RSV, and TK. Similarly, the reporter vector promoter is associated with at least one copy of an operator, as outlined herein.

5 In addition, it is possible to configure a retroviral vector to allow expression of bait genes or test genes after integration of a bait or test vector in target cells. For example, Tet-inducible retroviruses can be used to express bait or test genes (Hoffman et al., PNAS USA 93:5185 (1996)). Expression of this vector in cells is virtually undetectable in the presence of tetracycline or other active analogs. However, in the absence of Tet, expression is turned on to maximum within 48 hours after induction, with uniform increased expression of the whole population of cells that harbor the inducible retrovirus, indicating that expression is regulated uniformly within the infected cell population. A similar, related system uses a mutated Tet DNA-binding domain such that it bound DNA in the presence of Tet, and was removed in the absence of Tet. Either of these systems is suitable.

15 Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to nucleic acids which are to be expressed. "Operably linked" in this context means that the transcriptional and translational regulatory nucleic acid is positioned relative to any coding sequences in such a manner that transcription is initiated. Generally, this will mean that the promoter and transcriptional initiation or start sequences are 20 positioned 5' to the coding region. The transcriptional and translational regulatory nucleic acid will generally be appropriate to the host cell used, as will be appreciated by those in the art. Numerous types of appropriate expression vectors, and suitable regulatory sequences, are known in the art for a variety of host cells.

25 In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters, hybrid or synthetic promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention.

5

In general, the vectors of the present invention utilize two different types of promoters. In a preferred embodiment, the promoters on the bait and test vectors are constitutive, and drive the expression of the fusion proteins and selection genes, if applicable, at a high level. However, it is possible to utilize inducible promoters for the fusion constructs and selection genes, if necessary.

1000 1500

The test vector comprises a selection gene. Selection genes allow the selection of transformed host cells containing the vector, and particularly in the case of mammalian cells, ensures the stability of the vector, since cells which do not contain the vector will generally die. Selection genes are well known in the art and will vary with the host cell used. Suitable selection genes include, but are not limited to, neomycin, blastocidin, bleomycin, puromycin, hygromycin, and other drug resistance genes, as well as genes required for growth on certain media, including, but not limited to, His and Lev or His and Trp. In some cases, for example when using retroviral vectors, the requirement for selection genes is lessened due to the high transformation efficiencies which can be achieved. Accordingly, selection genes need not be used in retroviral constructs, although they can be. In addition, when retroviral vectors are used, the test vectors may also contain detectable genes as are described herein rather than selection genes; it may be desirable to verify that the vector is present in the cell, but not require selective pressure for maintenance.

25

In addition to the selection gene, the test vector comprises a fusion gene comprising a first sequence encoding a transcriptional activation domain, and a second sequence encoding a test protein. By "fusion gene" or "fusion construct" herein is meant nucleic acid that comprises at least two functionally distinct sequences; i.e. generally sequences from two different genes.

30

herein may be DNA, for example when extrachromosomal plasmids are the vectors, or RNA, for example when retroviral vectors are used. Generally, the sequences are directly linked together without any linking sequences, although in some embodiments linkers such as restriction endonuclease cloning sites or linkers encoding flexible amino acids such as glycine and serine linkers such as are known in the art are used. In a preferred embodiment, the first fusion gene comprises a first sequence encoding a transcriptional activation domain. By "transcriptional activator domain" herein is meant a proteinaceous domain which is able to activate transcription.

10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
9235
9240
9245
9250
9255
9260
9265
9270
9275
9280
9285
9290
9295
9300
9305
9310
9315
9320
9325
9330
9335
9340
9345
9350
9355
9360
9365
9370
9375
9380
9385
9390
9395
9400
9405
9410
9415
9420
9425
9430
9435
9440
9445
9450
9455
9460
9465
9470
9475
9480
9485
9490
9495
9500
9505
9510
9515
9520
9525
9530
9535
9540
9545
9550
9555
9560
9565
9570
9575
9580
9585
9590
9595
9600
9605
9610
9615
9620
9625
9630
9635
9640
9645
9650
9655
9660
9665
9670
9675
9680
9685
9690
9695
9700
9705
9710
9715
9720
9725
9730
9735
9740
9745
9750
9755
9760
9765
9770
9775
9780
9785
9790
9795
9800
9805
9810
9815
9820
9825
9830
9835
9840
9845
9850
9855
9860
9865
9870
9875
9880
9885
9890
9895
9900
9905
9910
9915
9920
9925
9930
9935
9940
9945
9950
9955
9960
9965
9970
9975
9980
9985
9990
9995
10000
10005
10010
10015
10020
10025
10030
10035
10040
10045
10050
10055
10060
10065
10070
10075
10080
10085
10090
10095
10100
10105
10110
10115
10120
10125
10130
10135
10140
10145
10150
10155
10160
10165
10170
10175
10180
10185
10190
10195
10200
10205
10210
10215
10220
10225
10230
10235
10240
10245
10250
10255
10260
10265
10270
102

into test vectors. The genomic library may be a complete library, or it may be fractionated or enriched as will be appreciated by those in the art.

5 In a preferred embodiment, the test protein sequences are derived from cDNA libraries. A cDNA library from any number of different cells may be used, and cloned into test vectors. As above, the cDNA library may be a complete library, or it may be fractionated or enriched in a number of ways.

10 In a preferred embodiment, the test protein sequences are random sequences. Generally, these will be generated from chemically synthesized oligonucleotides. Generally, random test proteins range in size from about 2 amino acids to about 100 amino acids, with from about 10 to about 50 amino acids being preferred. Fully random or "biased" random proteins may be used; that is, some positions within the sequence are either held constant or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or 15 amino acid residues are randomized within a defined class, for example, of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., for zinc fingers, SH-2 domains, stem loop structures, or to purines, or to reduce the chance of creation of a stop codon, etc.

20 The compositions of the invention also include reporter vectors. Generally, the test and reporter vectors are distinct, although as will be appreciated by those in the art, one or two independent vectors may be used. The reporter vectors comprise a first detectable or reporter gene and all or part of the switch ϵ sequence, which functions as an operator site. That is, 25 upon binding of a test protein to the switch ϵ sequence (i.e. a protein-nucleic acid interaction), the transcriptional activator domain of the fusion protein will activate transcription and cause expression of the selectable or detectable gene(s). Thus, in this embodiment, the test protein functions essentially as a candidate agent.

In a preferred embodiment, the compositions are introduced into host cells to screen for 5 protein-nucleic acid interactions. By "introduced into" or grammatical equivalents herein is meant that the nucleic acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type and the composition of the vector. Exemplary methods include CaPO₄ precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. The vectors may stably integrate into the genome of the host cell (for example, with retroviral introduction for mammalian cells, outlined herein), or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.).

10
15
20
25
30

150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300

20

25

30

The vectors can be introduced simultaneously, or sequentially in any order. In a preferred embodiment, host cells containing the reporter construct are generated first, and preferably the reporter vector is integrated into the genome of the host cell, for example, using a retroviral reporter vector. Once the components of the system are in the host cell, the cell is subjected to conditions under which the selectable markers and fusion proteins are expressed. If a test protein has sufficient affinity to the switch ϵ region to activate transcription, the detectable protein is produced, and cells containing these proteins will survive drug selection and can be detected as outlined above. The detectable protein will be produced at a measurably higher level than in the absence of a protein-nucleic acid interaction. Thus the determination of a protein-nucleic acid interaction is generally done on the basis of the presence or absence of the detectable gene(s).

In a preferred embodiment, once a cell with an altered phenotype is detected, the cell is isolated from the plurality which do not have altered phenotypes. This may be done in any number of ways, as is known in the art, and will in some instances depend on the assay or screen. Suitable isolation techniques include, but are not limited to, drug selection, FACS, lysis selection using complement, cell cloning, scanning by Fluorimager, expression of a "survival" protein, induced expression of a cell surface protein or other molecule that can be rendered fluorescent or taggable for physical isolation; expression of an enzyme that changes a non-fluorescent molecule to a fluorescent one; overgrowth against a background of no or

slow growth; death of cells and isolation of DNA or other cell vitality indicator dyes; changes in fluorescent characteristics, etc. The preferred isolation techniques are drug selection and FACS based on the expression of the detectable gene, with a preferred embodiment utilizing both simultaneously.

5

Once a cell with a protein-nucleic acid interaction is detected and isolated, it is generally desirable to identify the test protein. In a preferred embodiment, the test protein nucleic acid and/or the test protein is isolated from the positive cell. This may be done in a number of ways. In a preferred embodiment, primers complementary to DNA regions common to the vector, or to specific components of the library such as a rescue sequence, are used to "rescue" the unique test sequence. Alternatively, the test protein is isolated using a rescue sequence. Thus, for example, rescue sequences comprising epitope tags or purification sequences may be used to pull out the test protein, using immunoprecipitation or affinity columns. Alternatively, the test protein may be detected using mass spectroscopy.

15

Once a bioactive agent is identified, a number of things may be done. In a preferred embodiment, the characterization of the bioactive agent is done. This will proceed as will be appreciated by those in the art, and generally includes an analysis of the structure, identity, binding affinity and function of the agent. Depending on the type of agent, this may proceed in a number of ways. In a preferred embodiment, for example when the candidate agents have been introduced intracellularly using nucleic acid constructs, the candidate nucleic acid and/or the bioactive agent is isolated from the cells. This may be done in a number of ways. In a preferred embodiment, primers complementary to DNA regions common to the retroviral constructs, or to specific components of the library such as a rescue sequence, defined above, are used to "rescue" the unique random sequence. Alternatively, the bioactive agent is isolated using a rescue sequence. Thus, for example, rescue sequences comprising epitope tags or purification sequences may be used to pull out the bioactive agent, using immunoprecipitation or affinity columns. Alternatively, the peptide may be detected using mass spectroscopy.

30

Once rescued, the sequence of the bioactive agent and/or bioactive nucleic acid is determined. Similarly, candidate agents from other chemical classes can be identified and characterized, for example through the use of mass spectroscopy. This information can then be used in a number of ways.

5

In a preferred embodiment, the bioactive agent is resynthesized and reintroduced into the target cells, to verify the effect. This may be done using retroviruses, or alternatively using fusions to the HIV-1 Tat protein, and analogs and related proteins, which allows very high uptake into target cells. See for example, Fawell et al., PNAS USA 91:664 (1994); Frankel et al., Cell 55:1189 (1988); Savion et al., J. Biol. Chem. 256:1149 (1981); Derossi et al., J. Biol. Chem. 269:10444 (1994); and Baldin et al., EMBO J. 9:1511 (1990), all of which are incorporated by reference. Other techniques known in the art may be used as well.

10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
9235
9240
9245
9250
9255
9260
9265
9270
9275
9280
9285
9290
9295
9300
9305
9310
9315
9320
9325
9330
9335
9340
9345
9350
9355
9360
9365
9370
9375
9380
9385
9390
9395
9400
9405
9410
9415
9420
9425
9430
9435
9440
9445
9450
9455
9460
9465
9470
9475
9480
9485
9490
9495
9500
9505
9510
9515
9520
9525
9530
9535
9540
9545
9550
9555
9560
9565
9570
9575
9580
9585
9590
9595
9600
9605
9610
9615
9620
9625
9630
9635
9640
9645
9650
9655
9660
9665
9670
9675
9680
9685
9690
9695
9700
9705
9710
9715
9720
9725
9730
9735
9740
9745
9750
9755
9760
9765
9770
9775
9780
9785
9790
9795
9800
9805
9810
9815
9820
9825
9830
9835
9840
9845
9850
9855
9860
9865
9870
9875
9880
9885
9890
9895
9900
9905
9910
9915
9920
9925
9930
9935
9940
9945
9950
9955
9960
9965
9970
9975
9980
9985
9990
9995
10000
10005
10010
10015
10020
10025
10030
10035
10040
10045
10050
10055
10060
10065
10070
10075
10080
10085
10090
10095
10100
10105
10110
10115
10120
10125
10130
101

compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt.%.

The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, 5 pills, suppositories, capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for oral and topical use can be used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing agents, wetting and emulsifying agents, salts for varying the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents.

10 The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope 15.0 of this invention, but rather are presented for illustrative purposes. All references cited herein are incorporated by reference in their entirety.

EXAMPLES

Example 1

Construction of ϵ germline GFP/BFP knock-in cell lines

20 Three different IgM $^+$, EBV $^-$ human B cells lines (CA-46, MC116, DND39, Figure 4) that produce ϵ germline transcripts in the presence of IL-4 will be transfected with a germline ϵ GFP or BFP knock-in construct (Figures 5B and 5C) and induced with IL-4. The cells will 25 then be sorted by FACS for the appropriate reporter expression, GFP or BFP. Background (i.e. random integration) should be low since the construct must integrate downstream of an IL-4 inducible region in order to be activated. Homologous recombination of the reporter construct will be confirmed in fluorescent clones by genomic PCR using primers located within and immediately flanking the construct. For double knockouts, both GFP and BFP 30 constructs will be transfected and cells sorted for expression of both reporters.

It is possible that activation with IL-4 to identify homologous recombined clones will result in events that move beyond the first phase of ϵ switching, thus making the clones unusable for a screen identifying blockers of this first step. For this case, we have designed a more traditional construct containing an SV40 promoter-driven neomycin resistance gene which is flanked by loxP sites and inserted in the intron between the first and second ϵ constant coding exons (Figure 5D). In addition, attached at the 3' end of the long arm is a BFP reporter gene driven by a constitutive promoter. B cell clones transfected with this construct will be selected for integration by culturing them in the presence of G418. The surviving cells lacking BFP will be sorted by FACS (the BFP at the 3' end will be preferentially deleted during the homologous recombination event). The remaining clones will be assessed for homologous recombination by PCR. Clones containing homologous recombined constructs will be exposed to the *cre* recombinase protein to mediate excision of the SV40 promoter/neomycin resistance gene in order to eliminate promoter interference and potential ϵ promoter shutdown. Excision of the SV40 promoter/neomycin resistance gene fragment will be verified by subdividing clones into parent and daughter pools and re-selecting the latter pool in G418. The parental cells corresponding to G418 sensitive daughter cells will be subdivided again and tested for IL-4 inducible GFP expression. Parental stocks of the most inducible clones will be used for subsequent peptide screening. Production of the knock-in cell line using this approach would provide a continuous source of IL-4 inducible cells and would circumvent any down-regulation associated with IL-4 pre-treatment.

Example 2

Creation and screening of candidate bioactive agents in knock-in cell lines

A candidate bioactive agent library, in this case a peptide library, will be packaged into infectious viral particles as outlined below. A preferred library is a mixture of random peptide sequences with and without a nuclear localization sequence (NLS) upstream of a reporter gene to identify infected cells and relative peptide expression (see Figure 6).

Each screen will start with production of the primary retrovirus peptide library, as is generally shown in Figure 7. This is generally done as outlined in PCT US97/01019 and PCT US97/01048, both of which are expressly incorporated by reference. In general, this is done as follows. On day 1, the Phoenix cells are seeded in 10 cm plates at 5 X 106 cells in 6 ml (DMEM + 10% FBS + Pen/Strep) per plate the day before transfection. Day 2: allow all reagents to reach room temperature 30 min. before starting. Add 50 mM chloroquine at 8 μ l/plate (50 μ M final) before preparing the transfection solution. Mix CaPO₄ reagents in 15ml polypropylene tube: per plate: 10 μ g DNA, 122 μ l 2M CaCl₂, 876 μ l H₂O, 1.0ml 2X HBS. Add 2X HBS and depress the expulsion button completely to bubble air through the mix for 10 secs. Immediately add mixture gently dropwise to plate. Incubate 3-8 hours. Remove medium and replace with 6.0 ml DMEM-medium. Day 3: Change medium again to 6.0 mls of medium optimal for the cells to be infected. Move to 32°C either in the morning or afternoon depending on the Phoenix cell confluence and whether you will infect at 48 or 72 hrs after transfection. Day 4 or 5: Collect virus supernatant from transfected plates (6.0 ml) into 50 ml tubes and add protamine sulfate to a final concentration of 5 μ g/ml. Pass through a 0.45 μ m filter. Count target cells and distribute 10⁷ cells per 10 cm plate transfected to 50 ml tubes and pellet 5 min. Resuspend each pellet of target cells in virus supernatant and transfer to a 6 well plate at 1.0-1.2 ml per well. Seal plate with parafilm and centrifuge at RT for 30-90 min. at 2500 RPM. Remove parafilm and incubate plate over night at 37°C. Day 5: Collect and pellet each well of target cells. Resuspend in 3 ml medium and transfer back to the same 6well plate. Infection can be repeated by refeeding the Phoenix cells with 6ml fresh medium and reinfecting the same cells again up to 3 times to increase % of cells infected (for instance at 48, 56, and 72 hours). Day 7 or Day 8: At 48 to 72 hrs. post infection, target cells are ready to analyze for expression.

This primary library will be used to infect at least 10⁹ knock-in cells. After infection, the cells will be stimulated with IL-4 and two days later, peptide-containing cells (identified by the fluorescent reporter) that are negative for the knock-in reporter (i.e. where there is ϵ promoter inhibition) will be sorted by FACS. This enriched, knock-in reporter negative population will be subjected to RT-PCR to amplify the integrated peptide sequences. The

PCR material will be used to construct a new "enriched" retrovirus peptide library to initiate the next screening round.

It will take approximately 5-7 rounds of enrichment to identify individual sequences capable of inhibiting the germline ϵ promoter, as outlined below using an iterative screening equation.

V

R =
$$\frac{v}{\epsilon^p + (Q + \sum_{i=0}^{\infty} \beta (1 + \epsilon)^p) + \gamma}$$

The above equation mathematically models screening efficiency and provides a guideline for monitoring enrichment for inhibitory peptides. R = ratio of true positive cells over the total number of cells screened per round of selection; v = frequency of true positive cells (ie. # of cells expressing peptide inhibitors of IgE switch/synthesis); ϵ = frequency of non-heritable false-positive cells (ie. # of cells in which IgE switch/synthesis is inhibited due to stimulation/screening inefficiencies, but are IgE positive in subsequent selection rounds); p = number of rounds of selection/enrichment applied to library screen; Q = initial frequency of cells with an heritable false-positive phenotype (ie. dominant-negative somatic mutation in cells that prevent IgE switch/synthesis); β = frequency of false-positives incurred by or during the selection/enrichment process.

Since we amplify enriched peptides by RT-PCR after each selection round, the equation can be simplified to

$$R = \frac{v}{\epsilon^p + Q + \gamma}$$

By plugging in empirically-derived or estimated values for the variables, an estimate of how many selection rounds must be applied to a library before enrichment for IgE inhibitory peptide becomes apparent.

5 For the purposes of our screens, we engineer and select reporter cell lines in which the values of ϵ and Q are low to minimize the number of screening rounds necessary to observe rare positive peptide "hits".

For example, IL-4 treatment upregulates the IgE switch reporter in 97% of cells, therefore $\epsilon = 0.03$. Of the uninduced cells, a second round of stimulation indicates that less than .01% of the starting population contain heritable false positives, therefore $Q < 0.0001$. A conservative estimate of IgE inhibitory peptides in the starting population is $1/10^8$, therefore $\nu = 10^{-8}$. Solving the equation for the number of selection rounds required to enrich to 50% true positive hits...

$$0.5 = \frac{10^{-3}}{(0.03)^p + 10^{-3} + 10^{-3}} \rightarrow p = 5 \text{ rounds}$$

15 The most important factor that influences the number of enrichment rounds necessary to identify individual peptide hits is the ratio between the real positive peptide hits in the original library and the heritable false positives. The frequency of real positive peptide hits is dependent upon the qualitative ability of the peptide to access and, in the correct conformation, bind to regulatory domains on proteins in the pathway of interest. Thus, 20 preferably, multiple scaffolding structures are used for presentation of random peptide surfaces and also different localization sequences fused to those peptide structures. Enrichment of real positive peptides becomes less efficient with false positive rates above 2%. For this reason, great emphasis is placed on developing robust reporter constructs and cell lines.

Uneven RT-PCR amplification may decrease overall amplification of real peptides hits from one round to another. This is overcome by additional rounds of library enrichment and is why RT-PCR amplification is carefully monitored after each round of screening.

5

Example 3

Screening for inhibitors of IgE secretion in cells that have already switched

10 After B cells have switched to production of IgE, there are several factors that determine when they will secrete IgE. By screening for peptide inhibitors of surface IgE expression, proteins that regulate IgE transcription, translation, assembly and trafficking may be identified.

15 The IgE⁺ cell line, U266, expresses IgE on the surface and also secretes IgE. Antibodies against surface IgE heavy and light chains have been obtained and both are used to fluorescently mark IgE positive cells. The U266 line is consistently greater than 98.5% positive for membrane IgE.

20 Peptide library screening and target identification: The peptide library and enrichment protocols identical to those described in Example 2. As well, peptide hit validation and corresponding target protein identification will be performed as described in Example 2.

25 Development of an ϵ -heavy chain GFP/BFP knock-in cell line derivative of U266: The cytoplasmic tail of the ϵ -heavy chain in U266 cells will be engineered by homologous recombination to encode a GFP/BFP reporter as shown in Figure 8. This will produce a cell line that is fluorescent when ϵ -heavy chains are produced. The GFP will be attached to the secretory exon to label secretory IgE heavy chains green. The BFP will be attached to the M2 exon to label membrane IgE heavy chains blue. This will allow discrimination of mRNA processing and translation between secretory versus membrane ϵ -heavy chain transcripts.

The construct will contain an SV40 promoter-driven neomycin resistance gene which is flanked by loxP sites and inserted in the intron between the CH3 and CH4 exons (Figure 8). In addition, the HSV-TK gene will be cloned 3' of the longer homologous sequence region. U266 cells transfected with this construct will be selected for integration by culturing them in 5 the presence of G418. The surviving cells will be cultured in ganciclovir to select against cells containing the HSV-TK gene (the HSV-TK gene at the 3' end will be deleted during the desired homologous recombination event). The remaining clones will be assessed for homologous recombination by PCR. Clones containing homologously-recombined constructs will be transfected with *cre* to mediate excision of the SV40 promoter/neomycin resistance gene in order to eliminate promoter interference. Excision will be verified by subdividing clones into parent and daughter pools and re-selecting the latter pool in G418. The parental cells corresponding to G418 sensitive daughter cells will be subdivided again and tested for 10 GFP and BFP expression. Parental stocks of the most inducible clones will be used for subsequent screening.

15

Example 4

Development of an ϵ promoter GFP reporter cell line

The induction of the ϵ promoter in response to IL-4/13 is the first recognizable step necessary 20 for the switch to IgE. Blocking activation of this promoter should prevent B cells from switching to IgE. Inhibitors are predicted to interfere with IL-4/13 signaling as well as nuclear transcription of the ϵ germline gene.

25 Three IgM⁺, EBV⁻ human B cells lines (CA-46, MC116, and DND39; see Figure 4) that produce ϵ germline transcripts in the presence of IL-4 will be infected with the following construct: a retroviral vector containing an IL-4 responsive 600 bp fragment of the ϵ promoter in the reverse orientation followed by a splice site, GFP encoding sequence and a poly-adenylation sequence (Figure 10). Briefly, cells will be infected with the reporter construct and induced with IL-4. The cells will then be sorted by FACS for GFP reporter 30 expression. The IL-4 will be removed and the cells will be sorted for the absence of reporter

fluorescence. From these sorts, several clones will be established that turn on the reporter in the presence of IL-4, indicating activation of the germline ϵ promoter.

Example 5

5 Screening of candidate agents using reporter cell line

The cell line of Example 4 is infected infected with a peptide library as described above. The peptide library is packaged into infectious viral particles (see Figure 7). The library is a mixture of random peptide sequences with and without a nuclear localization sequence (NLS) upstream of a reporter gene to identify infected cells and relative peptide expression (Figure 6).

Each screen will start with production of the primary retrovirus peptide library. This primary library will be used to infect at least 10^9 ϵ promoter reporter cells. After infection, the cells will be stimulated with IL-4 and two days later, the FACS will sort peptide-containing, reporter negative cells (i.e. where there is ϵ promoter inhibition). This enriched, reporter negative population will be subjected to RT-PCR to amplify the integrated peptide sequences. The PCR material will be used to construct a new "enriched" retrovirus peptide library to initiate the next screening round.

20 It will take approximately 5-7 rounds of enrichment to identify individual sequences capable of inhibiting the germline ϵ promoter (see discussion above regarding the statistics associated with enrichment). The most important factor that influences the number of enrichment rounds necessary to identify individual peptide hits is the ratio between real positive peptide hits in the original library and heritable false positives. The frequency of real positive peptide hits is dependent upon the qualitative ability of the peptide to get to and, in the correct conformation, bind to the regulatory domains on proteins in the pathway of interest. This is why we use multiple scaffolding structures for presentation of random peptide surfaces and also different localization sequences fused to those peptide structures (Appendix B).

25 30 Enrichment of real positive peptides becomes less efficient with false positive rates above

2%. For this reason, great effort is placed in developing robust reporter constructs and cell lines.

Once enrichment is achieved and individual peptide sequences are shown to effect inhibition
5 of ϵ promoter activation in an independent assay, they will be introduced into a standard set of secondary and orthogonal assays. Many of these assays will be performed in primary B cells to test the specificity and physiologic characteristics of the peptide inhibitor.

Example 6

Generation of an ϵ promoter survival cell line.

0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150

Three different IgM⁺, EBV⁻ human B cell lines that produce ϵ germline transcripts in the presence of IL-4 will be infected with a survival construct carrying a death gene and a drug selectable marker (Figure 10). Briefly, the retroviral construct consists of the 600 bp IL-4 inducible ϵ promoter downstream of a self-inactivating (SIN) LTR, followed by a chimeric FAS receptor (FASr), the self-cleaving peptide 2a and, lastly, the drug-selectable puromycin resistance gene. The chimeric receptor is composed of the mouse FASr external domain and the human FASr transmembrane and cytoplasmic domains. A mouse specific anti-FASr antibody can be used which will bind only activated FASr produced by the survival construct.
20 The 2a self-cleaving peptide allows equimolar amounts of the chimeric FASr and puromycin to be produced in the cell.

20

25

IgM⁺ B cell lines infected with this construct in the presence of IL-4 will produce CD95, as well as puromycin resistance. Upon drug selection with puromycin, only cells containing IL-4 activated ϵ promoters will survive. The remaining cells are infected with the peptide libraries and, when cultured in the presence of IL-4 and anti-FAS (α CD95) monoclonal antibodies, will express the chimeric FAS receptor and apoptose unless their ϵ promoter has been blocked by a library peptide.

10 If problems arise due to over-expression of the chimeric FASr resulting in self-activation, other external domains will be used. We have already engineered a chimeric FASr containing the murine CD8 external domain as an alternative (Figure 10). If overexpression of the chimeric FASr results in self-activation, we have designed an alternative strategy in which the
5 proposed construct contains the GFP gene in lieu of the puromycin resistance gene (Figure 10). Due to the mild transcriptional leakiness inherent to all SIN retroviral vectors, a small percentage of IgM⁺ B cell clones infected with this construct will express low, detectable levels of GFP. These cells can be single-cell cloned by FACS, split into parent and daughter pools and tested for IL-4 inducible FASr expression-dependent apoptosis. Parent stocks of the most efficiently killed daughter cells will provide a continuous cell source for subsequent peptide screening assays. In addition, FASr ligation can be used to potentiate cell death and thus diminish background cell survival.

15 Additionally, IL-4 stimulation has been reported to diminish FAS-induced apoptosis in certain B-cell lines. To circumvent this potential difficulty, common suicide genes including Herpes Simplex Virus Thymidine Kinase (HSV-TK) or human cytochrome P450 2B1 in conjunction with ganciclovir or cyclophosphamide treatment, respectively, can replace FASr-mediated death (Figure 10). Alternatively, cell cycle arrest genes such as p21 can be used in place of toxic gene products (Figure 10). In this way, cells expressing peptides which
20 prevent IL-4 induced overexpression of p21 will have a selective growth advantage and will quickly dominate the culture.

Example 7

Screening in ϵ promoter survival cells

25 Using a peptide library generated as outlined above, the IgM⁺ B cell lines described in Example 6 are infected with the survival construct. Leaky cells (constitutive expression of the ϵ promoter) will be removed by incubation with the anti-mouse FASr antibody. Next, the cells are incubated in the presence of the inducer, IL-4, and the drug selection compound,
30 puromycin. Cells that contain a construct that is inducible by IL-4 will be resistant and

survive. This produces a population with an exogenous ϵ promoter that is IL-4 inducible. The peptide library is introduced into these cells and two days later they are induced with IL-4 in the presence of anti-mouse FASr monoclonal antibody. Cells carrying peptides that inhibit induction of the engineered ϵ promoter fragment will not produce the chimeric FASr and will 5 survive. After the survivors grow out (approximately 1 week), they will again be subjected to IL-4 and the anti-FASr treatment. The genes encoding the peptides responsible for the survivors will be rescued by RT-PCR and used to generate an enriched retroviral library. The identification of individual inhibitory peptides should occur in only 3-4 rounds since the false positive background for survival screens is lower than for FACS-based screening. Once 10 enrichment is achieved and individual peptide sequences are independently shown to inhibit ϵ promoter activation, these sequences will be introduced into a standard set of secondary and orthogonal assays.

Example 8

One-hybrid screens for identification of proteins that bind to switch ϵ region

20 Recombinase proteins that bind to the S ϵ region mediate the DNA rearrangement that generates a functional ϵ heavy chain. They may be specific for ϵ switching cells or may bind to other proteins that target them specifically to the S ϵ region. Breakpoints in the recombination of the switch ϵ region to the switch μ region occur in a limited area of the switch ϵ region. Two stretches of the switch ϵ region spanning the majority of breakpoints will be used as bait in a one-hybrid screen (Figure 2b). The cDNA libraries to be used are derived from the IgE positive cell line U266 (the assumption here is that the U266 line still 25 contains the switch recombinase; certainly, the recombinase is turned off in plasma cells) and from human peripheral blood lymphocytes stimulated *in vitro* to switch with a high frequency to IgE.

30 The screening is summarized in Figure 3. The methods are as follows: Two stretches of the switch ϵ region were cloned (Figure 2A) into EcoR I/Xba I sites of pHIS1-1 (Clontech) to

construct a HIS reporter vector pIgE-HIS. In this construct, HIS expression is under the control of a minimal promoter and proteins binding to the switch ϵ region. Similarly, a second LacZ reporter is constructed by inserting two stretches of switch ϵ region into the EcoR I/Xho I sites of pLacZi to construct pIgE-LacZ.

5

The pIgE-HIS was linearized at an Afl II site and integrated into yeast strain YM4271 (MAT α , ura3-52, his3-200, ade2-101, lys2-801, leu2-3, 112, trp1-901, tyr1-501, gal4-Δ512, gal80-Δ538, ade5::hisG) to construct the first yeast reporter strain YIgE-HIS. SD-H plates were used to select for integrated reporters. The yeast strain YIgE-HIS was tested on SD-H+3AT plates to determine the optimal concentration of 3AT to suppress basal level HIS expression from the minimal promoter.

1538 1539

The pIgE-LacZ plasmid was linearized at an Nco I site and integrated into the yeast strain YIgE-HIS to construct a dual reporter strain YIgE-HL. SD-U plates were used to select for cells with dual reporters integrated. The dual reporter strain will be used for transformation by the U266 cDNA library (it is assumed that the U266 line still contains the switch recombinase) and the IgE switching PBL cDNA library. At least 20 million transformants from each library will be screened on SD-LH+3AT plates. Clones that can grow up and turn blue on SD-LH+3AT plates will be grown up on SD-LH medium and further identified by

20